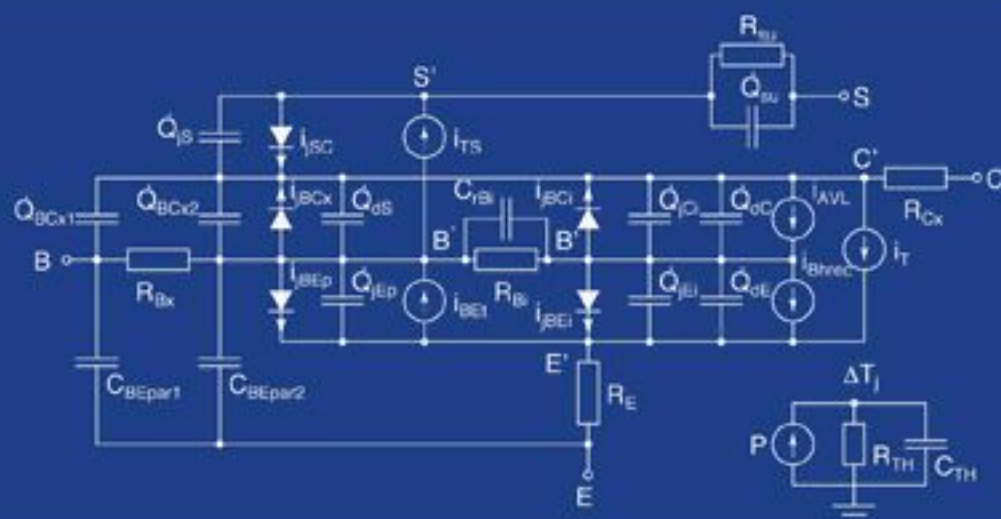


**ASSET**

International Series on Advances in Solid State Electronics and Technology  
Founding Editor: Chih-Tang Sah

# COMPACT HIERARCHICAL BIPOLAR TRANSISTOR MODELING WITH HICUM



**Michael Schröter**  
**Anjan Chakravorty**

**World Scientific**

# **COMPACT HIERARCHICAL BIPOLAR TRANSISTOR MODELING WITH HICUM**

**ASSET**

**International Series on Advances in Solid State Electronics and Technology**

**Founding Editor: Chih-Tang Sah**

# **COMPACT HIERARCHICAL BIPOLAR TRANSISTOR MODELING WITH HICUM**

**Michael Schröter**

University of Technology Dresden, Germany  
University of California San Diego, CA, USA

**Anjan Chakravorty**

Indian Institute of Technology Madras, India

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

**International Series on Advances in Solid State Electronics and Technology  
(ASSET)**

*Founding Editor:* Chih-Tang Sah

---

*Published:*

Modern Semiconductor Quantum Physics

*by* Li Ming-Fu

Ionizing Radiation Effects in MOS Oxides

*by* Timothy R. Oldham

MOSFET Modeling for VLSI Simulation: Theory and Practice

*by* Narain Arora

MOSFET Modeling for Circuit Analysis and Design

*by* Carlos Galup-Montoro and Márcio Cherem Schneider

The Physics and Modeling of MOSFETs: Surface-Potential Model HiSIM

*by* Mitiko Miura-Mattausch, Hans Jürgen Mattausch & Tatsuya Ezaki

Invention of the Integrated Circuits: Untold Important Facts

*by* Arjun N Saxena

Electromigration in ULSI Interconnections

*by* Cher Ming Tan

Compact Hierarchical Bipolar Transistor Modeling with HICUM

*by* Michael Schroter and Anjan Chakravorty

*Forthcoming:*

BSIM4: Theory and Engineering of MOSFET Modeling for IC Simulation

*by* Weidong Liu and Chenming Hu

*Published by*

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

*USA office:* 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

*UK office:* 57 Shelton Street, Covent Garden, London WC2H 9HE

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**COMPACT HIERARCHICAL BIPOLAR TRANSISTOR MODELING WITH HICUM**  
**International Series on Advances in Solid State Electronics and Technology**

Copyright © 2010 by World Scientific Publishing Co. Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-4273-21-3

ISBN-10 981-4273-21-X

**Disclaimer:** This book was prepared by the authors. Neither the Publisher nor its Series Editor thereof, nor any of their employees, assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information. The contents, views, and opinions of the authors expressed herein do not necessarily state or reflect those of the Publisher, its Series Editor, and their employees.

Printed in Singapore.

***To my wife Ursula and my parents for their  
patience, understanding and support***

***Michael Schröter***

***To my wife and my parents ....***

***Anjan Chakravorty***

This page is intentionally left blank

# Foreword

This monograph is the sixth book in this series on modeling of integrated-circuit devices. The purpose of this series is to provide archival references, described by the model originators or authorities, on the electron devices that are interconnected on a small, about a square centimeter or smaller, silicon semiconductor integrated circuit chip. It has been my intention to invite the experts to cover all the devices of an integrated circuit. These include the active devices: field-effect transistors with insulated, MOS, p/n-junction and metal/semiconductor-junction gates, and bipolar-junction transistors; the passive devices: diodes, resistors, capacitors and inductors; and the devices which interface the electronics with the ambient and the users: electromagnetic and optical radiators-receivers and their antennas, the transducers of temperature, pressure, strain, stress, flow velocities, and force accelerations, and the sensors of biological and chemical molecules; and certainly the new electrical-mechanical and quantum nanometer-dimension devices when proven in applications.

The monograph series idea came about when I was looking into the literature for references to write the keynote address for the fourth annual International Workshop on Compact Modeling, to be presented on May 10, 2005. I was invited by the Workshop Founder, Professor Xing Zhou of Nanyang Technology University (NTU), and his workshop program committee members. We decided to review the history of MOS transistor compact modeling [1] in order to supplement a 10-author review presentation [2] on the state-of-the-art of MOS transistor compact models. The efforts of industrial-standard compact-transistor-modeling began in 1996 with the consensus first-generation BSIM of Professor Chenming Hu of UC Berkeley using my 1964-1965 threshold-voltage model [3,4]. After a decade, it was to advance to an industrial consensus second generation model in 2006, using my 1965-1966 Sah-Pao surface-potential model [4,5], with the surface potential as the parameter in order to continuously characterize the current ranges of saturation, linear, subthreshold, flat-band and cut-off, in the MOS transistor current-voltage characteristics, more accurately than the threshold-voltage model. HiSIM, PSP and BSIM models were the choices. It was to be the workshop's first keynote and it got me out of 40-year hibernation on MOS field-effect transistor modeling. A second purpose of this device modeling series was to provide state-of-the-art textbooks for graduate students and reference books for practicing engineers. These books will provide timely disseminations of detailed design methodologies and their underlying physics. To lead the technology which has been progressing at an ever faster pace, the design art must keep pace to provide the computer-aided-design of silicon semiconductor integrated circuits which contain hundreds, thousands, and now billions, approaching if not already exceeding a trillion, devices and transistors on one small ( $< \sim 1$  square centimeter) silicon die or silicon integrated circuit chip. It is also the objective of this monograph series to provide timely updates via Internet exchanges between the readers and authors, for immediate public dissemination, and



to provide update editions when enough materials are accumulated.

I am especially thankful to the invited authors of the four startup volumes (Narain, Carlos+Márco, Mitiko+Hans+Tatsuya, and Arjun), and of the later volumes (Cherming, Chenming+Weidong, Michael+Anjan of this volume, and Chinghsiang.). They concurred with my objectives and agreed to take up the chore of writing their books during their very busy schedules. Some have delays of one, two or even three years. Nevertheless, their monographs are still the archival records of the state of the art, and the world's authoritative contributions to the device modeling literature, because these authors are the creators, inventors and/or authorities of the models, and because the models are the industry-wide consensus models used by all circuit designers in the past and recent generations, and also of the future generations.

The present volume is a review of the design literature and compact model, HICUM, on bipolar junction and heterojunction bipolar transistors (BJT and HBT). They are used in the front end of portable devices such as cell phones in (~10 to 50+) Giga-Hertz frequency bands. They are especially important because of their power efficiency and large signal-to-noise ratio at the Giga-Hertz frequency ranges, which are not yet attained by the nanometer-dimension silicon MOS field-effect transistors.

The notations and terminologies used in this monograph follow the industrial practice used by the design and manufacturing engineers of the BJTs and HBTs in silicon and compound semiconductors. Most of the transistor notations and terminologies used here follow the 1949 Shockley article on the invention of the p/n junction transistor [6]. For example, the base layer thickness of the lateral (horizontal or x-axis) one-dimensional BJT was called the base layer width,  $W$ , by Shockley. It was indeed the geometric width of the thin 1-D base layer. However, the later and recent micrometer and nanometer BJTs and HBTs are 3-D and their width and length are in the plane perpendicular to the current, flowing through the thin nanometer-thickness base layer of the transistor. For another example, the correct quantum mechanical physics model of energy band and energy gap was used since the 1936 A. H. Wilson energy band and electron-hole theory [7] and the 1928 Felix Bloch and 1930 Leon Brillouin periodic structure theory that gave the frequency-wavenumber or omega-beta diagrams [8]. But, the 20<sup>th</sup> century engineers replaced them by the confusing and erroneous-physics word “bandgap”. The term is correct in English grammar to designate simultaneously both the energy band and the energy gap; the gap does not have a “band” of closely spaced energy levels. However, the engineers use the “bandgap” only to designate the energy gap, not both the energy gap and the energy band. Then they call the discontinuities in the energy band or energy gap in a heterojunction as “bandgap discontinuity”, resulting in two meanings of the word bandgap: band and gap, and band's gap or the gap of a band of closely spaced and allowed energy levels. For a third example, the device engineer's understanding of the difference between macroscopic and microscopic is

inconsistent with the standard and accepted usage of the physicists. Finally, most of the device engineers who develop the transistor theory are device “physicists” without the traditional industrial-academic engineering training in circuit theory. Thus, the physics and engineering distinction and mathematical difference between charge-controlled small-step-function-signal theory and the small-sinusoidal-signal theory are not recognized. And the IEEE notation standard of upper and lower cases of symbols and their superscripts and subscripts that distinguishes these differences are not followed because of the ignorance of the FORTRAN and modern computer language on upper and lower cases, and superscripts and subscripts, due to the insufficient word length and memory space, if not also the speed, of the earlier computers. To make progress, nevertheless, we in this book shall follow the popular practice rather than the unambiguous notations and correct physics. For further discussions on these choices and examples of mixed usage, see my college textbook published by WSPC in 1991, *Fundamentals of Solid-State Electronics* [9].

I would like to thank all the WSPC production staff members and copy editor Mr. Tjan Guangwei at Singapore, and the acquisition editor, VP Ms. Zhai Yubing in New Jersey, for their efforts and supports. Special thanks are due NTU Physics Professor Kok-Khoo Phua, the Founder and Chairman of WSPC, for his farsight, which is a proof of my thesis, the Evolutionary Intelligent Design. I also thank Dr. Jie Binbin, Professor of Physics of Xiamen University, for editorial assistance, and for his supports from President Zhu Chongshi, School of Physics Dean Wu Chenxu, and Department of Physics Chair Zhao Hong, all of Xiamen University, China.

Sah ChihTang (Chih-Tang Sah)

Department of Physics, Xiamen University, China.

CTSAH Associates, Florida, USA.

Seventh year on August 13, 2010 and advancing.

## References

- [1] Chih-Tang Sah and Bin B. Jie, “A History of MOS Transistor Compact Modeling,” Keynote, Technical Proceedings, Workshop on Compact Modeling, Anaheim, California. Abstract, pp.1-2; Full text, pp.349-390, May 8-12, 2005. <http://www.nsti.org/publ/Nanotech2005WCM/1429.pdf>
- [2] J. Watts, C. McAndrew, C. Enz, C. Galup-Montoro, G. Gildenblatt, C. Hu, R. van Langevelde, M. Miura-Mattausch, R. Rios and C-T Sah, “Advanced Compact Models for MOSFETs,” *ibid.*, pp.3-12.
- [3] Chih-Tang Sah, “Characteristics of the Metal-Oxide-Semiconductor Transistors,” *IEEE Transactions on Electron Devices*, 11(7), 324-345, July 1964.
- [4] Chih-Tang Sah and Henry C. Pao, “The Effects of Fixed Bulk Charge on the Characteristics of Metal Oxide-Semiconductor Transistors,” *IEEE Trans. Electron Devices*, 13(4), 410-415, April 1966.
- [5] Henry C. Pao and Chih-Tang Sah, “The Effects of Diffusion Current on the Characteristics of MOS Transistors,” *Solid-State Electronics*, 9(10), 927-938, October 1966.
- [6] William Shockley, “The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors,” *Bell System Technical Journal* 28, 435-489, July 1949.
- [7] A. H. Wilson, *The Theory of Metals*, Cambridge University Press, London, 1936. (2<sup>nd</sup> Edition 1958).
- [8] Harry Jones, *The Theory of Brillouin Zones and Electronic States in Crystals*, North-Holland, 1960.
- [9] Chih-Tang Sah, *Fundamentals of Solid-State Electronics*, 1001pp, World Scientific Publishing Company, Singapore, 1991. See Chapters 1, 2, and 3 for basic semiconductor materials and device physics, and the ~300-page Chapter 7 (pp.701-991) on all device-materials-circuit aspects of BJTs.

## Biographies

**Dr. Michael Schroter** received his electrical engineering Ph.D. in 1988 from the Ruhr-University Bochum, Germany. He was with Nortel, Ottawa, Canada, as Team Leader and Advisor until 1996 when he joined Rockwell (now Conexant), Newport Beach, California, where he managed the RF Device Modeling Group. Dr. Schroter has been a Full Professor at the University of Technology at Dresden (UTD), Germany since 1999, and a Research Professor at UC San Diego, USA. For several years, he was on the Technical Advisory Board of RFMAGIC (now Entropic Inc.), a communication system design company in San Diego, California. Dr. Schroter is a co-founder of XMOD Technologies in Bordeaux, France. He is the author of the industrial standard bipolar transistor compact model HICUM, the subject of this monograph. Since 2008, he has been the Technical Program Manager of DOTFIVE, a large European Research project on Half-THz SiGe HBT technology. He is presently on Leave of Absence from TUD to assume the position of Vice President of RF Engineering at RFNano, Newport Beach, California, where he is responsible for the device design of the first production carbon nanotube FET process technology.



**Dr. Anjan Chakravorty** received his B. Tech. degree from the University of Kalyani, India, in 1999, M. Tech. from the University of Calcutta, India, in 2001, and Ph.D. from Indian Institute of Technology Kharagpur in 2005. He was a Guest Scientist with the Modeling Group, Innovations for High Performance Microelectronics, IHP GmbH, Germany for nine months between 2003 and 2004. From 2005 to 2006, he was a Postdoctoral Fellow at Dresden University of Technology, Germany, where he worked on bipolar transistor model, HICUM. Since January 2007, he has been an Assistant Professor at the Department of Electrical Engineering, Indian Institute of Technology Madras. He has published in reputed international journals, mainly in the area of compact modeling. His current research interests include compact modeling of bipolar and MOS transistors, on-chip inductors and circuit design for RF applications.



## Preface

Why a book on bipolar transistor compact modeling? First, looking at existing books on bipolar transistors we felt that *compact modeling*, especially with emphasis on the *practical aspects* of model formulations for *advanced* technologies, was not receiving adequate attention. As a second point, during numerous discussions and meetings over the past couple of years, a growing number of HICUM users suggested to write this book, which therefore is an attempt to answer their many questions about the model through a comprehensive documentation. Third, HICUM is one of the standard compact models for bipolar transistors, which should provide additional justification. And yes, we know we are somewhat late but obviously not too late, since HICUM appears to be capturing an increasing number of users and since development of new and faster HBT technology still appears to be attractive enough to get funded. Finally, we should of course expect the above question to be asked by those potential readers who are not so familiar with bipolar technologies or HICUM or both. Therefore, a more detailed answer will be given in the Introduction as well as the last chapter on “Future Trends”.

This book is organized as follows. The Introduction in Chapter 1 discusses the relevance of bipolar transistor technology in the existing analog circuit applications and, based on selected examples, derives the most important requirements for compact bipolar transistor models. Within this perspective, a brief history of HICUM development is given. Chapter 2 starts with an overview on major bipolar technology generations and resulting typical transistor structures. Then a more detailed discussion is given on the various approaches of device modeling and on the specific requirements for compact modeling. The foundations for understanding the formulations and equivalent circuit employed in HICUM are laid down

in Chapters 3 to 7. While Chapter 3 still assumes a homojunction transistor, Chapter 4 extends the resulting theory to effects both related to the material composition in heterojunctions and generally found in advanced devices. In both chapters the one-dimensional case is treated to capture the intrinsic transistor operation. In contrast, Chapter 5 deals with the complete structure and related geometry scaling, starting with the internal transistor and related internal base resistance, then discussing perimeter injection effects, and finally closing with the consideration of three-dimensional effects. The impact of temperature changes resulting from both environment and self-heating on the model parameters and device characteristics is discussed in Chapter 6. An overview on the basic approach to derive analytical expressions representing carrier fluctuations (noise) in compact form is given in Chapter 7. Note that these chapters focus on *vertical npn* transistors. An extension to vertical *pnp* transistors is fairly straight-forward, but requires the addition of the EC elements of the *n*-well. The basic theory of lateral bipolar transistors is a subset of that in Chapter 3 for the intrinsic part but usually requires special geometry scaling and the addition of two parasitic transistor elements.

Equipped with the foundations above, the reader can then move on to the second part of this book, namely the compact model hierarchy. In Chapter 8 the complete equivalent circuit and associated set of equations for HICUM/L2 are constructed from the theoretical results of the previous chapters. Subsequently, Chapter 9 outlines a generic geometry-scalable parameter extraction sequence. Then, Chapter 10 first covers the derivation of the simplified HICUM/L0 version and finally discusses the various options for the distributed models that result in HICUM/L4. An overview on applications with comparisons to experimental data is given in Chapter 11. If possible (and not restricted by confidentiality), examples from independent users at industry and academia are shown instead of our own ones, since we believe that model developers will always be able to show excellent results. It is interesting to note, that in most cases we only hear from users if they encounter a problem with the model. However, it would help model development and documentation tremendously if users would also provide positive feedback, e.g., in form of comparisons between *circuit* simulations and measurements, and of directions regarding necessary and useful model extensions. The book is completed by

Chapter 12 with a detailed discussion of future trends and developments of HBT technology and modeling.

The development of a compact model such as HICUM has spun over more than 25 years at the time this book is published. Although most of the initial basic ideas proved to be adequate for the many technology generations that evolved during this time frame, the continuous improvement and extension towards the latest process generation have gradually changed the model and its complexity. This development has been accompanied by extensive experimental verifications as well as the adaptation of the necessary parameter extraction and device simulation infrastructure. Therefore, over the period of time mentioned above many individuals have contributed to reach the present state of the model. Below, we would like to thank these individuals.

First of all, the early work was initiated by Prof. Hans-Martin Rein at the Ruhr-University Bochum, Germany, whose insights into the workings of bipolar transistors and high-speed circuit design were instrumental in keeping the right balance between theory and application. Learning from his experience provided the first author (MS) with the foundations for his later work on bipolar transistors. Further contributions to model development came from Dr. Andreas Koldehoff, Dr. Martin Friedrich, Dr. Tzung-Yin Lee, Zhixin Yan, Dr. Hoang Hung Tran, and are still being provided by Mr. Steffen Lehmann, Mr. Andreas Pawlak, and Mr. Yves Zimmermann. Special thanks in this respect belong to the late Prof. David Jonathan Walkey, who during his first sabbatical in 1994 not only helped the first author (MS) to settle after his move to Ottawa, Canada, and became a good friend, but also was a frequent discussion partner who always took the time to thoroughly check many of the technical ideas.

The second author (AC) would like to thank Prof. Chinmoy Kumar Maiti of Indian Institute of Technology (IIT) Kharagpur for encouraging research on compact modeling of SiGe HBTs. While learning the subject, a new excitement came when IHP Microelectronics, Germany invited the second author (AC) as a guest scientist to work on modeling of some special SiGe:C HBT structures. Many individuals came into contact among whom Dr. Rene Scholz, Dr. Biswanath Senapati, and Dr. Samiran Halder continued as helpful friends. During the postdoc time of the second author (AC) with the first author (MS), AC got the unique opportunity to interact

with each and everyone at CEDIC. Meanwhile in 2007, the second author (AC) returned to India and joined IIT Madras, where the colleagues at Dept. of Electrical Engineering are found to be very much approachable for any academic help starting from extending simulation facility to encouraging academic discussion. Then came the golden jubilee celebration of IIT Madras and the book writing scheme as part of the celebration extended the second author (AC) a great help to complete this volume in time. Friends and family members were always there to mentally facilitate towards finishing this writing project.

The best compact model will not be successfully applied without working parameter extraction procedures and their proper execution. Therefore, we would like to thank Dr. Bertrand Ardouin, Mr. Didier Celi, Dr. Dominique Berger, Dr. Ramana Murti-Mallardi, Dr. Sadayuki Yoshitomi, Mr. Burkhard Stritzke, Mrs. Julia Krause, Mr. Hai Tran for their contributions. Also important in this respect is the in-kind support both in terms of hardware and software as well as financial support and internship opportunities for our students (enabling technology transfer). Here, we would like to thank Drs. Wolfgang Kraus, Mr. Winfried Rabe, Dr. Hans-Joachim Wassenner from TelefunkenSemi (formerly Atmel Germany); Mr. Didier Celi, Dr. Herve Jaouen, Dr. Andre Juge from STMicroelectronics (France); Dr. Marco Racanelli from Jazz Semi (USA); Dr. David Haramé from IBM (USA); Mr. Jörg Berkner and Dr. Karl Schön from Infineon (Germany); Dr. Peter Zampardi from Skyworks (USA); Dr. Franz Sischka and Dr. Rick Poore from Agilent (Germany, USA); Dr. Marek Mierzwinski from Tiburon (USA); Dr. Sadayuki Yoshitomi from Toshiba (Japan); Ms. Tracey Krakowski and Linda Smith from National Semiconductor (USA); Mr. Ahmed Ramadan and Dr. Phillipe Raynaud from MentorGraphics (USA, France); Dr. Jushan Xie from Cadence (USA); and all members of the CMC (USA). The first author (MS) also owes special thanks to Dr. Paulius Sakalas for his tireless effort of not only helping to build a world-class characterization lab but also for performing high-quality and specialized measurements used for model verification, and to Dr. Bertrand Ardouin who was “the right person at the right time” for setting up XMOD Technologies in 2003, which since then has helped the deployment of HICUM tremendously through the commercialization of reliable parameter



extraction software and through providing extraction services for foundry design kits.

We appreciate the continuous support and feedback of many individuals regarding model implementation (including solving Verilog-A related issues) and testing in circuit simulators: Dr. Y.-C. Yuan, Mrs. Rosana Perez from Agilent (USA); Mr. Mohamed Selim, Mr. Thierry Fafournoux, Dr. Joel Besnard from MentorGraphics (France); Dr. Jean-Paul Malzac from Silvaco (France); Dr. Adam Divergilio from Tektronix (USA); Mr. Kai-Eric Moebus (now at CEDIC) and Dr. Alberto Piazza from AWR (USA); Dr. Ilja Yusim, Dr. M. Yakupov from Cadence (USA); Mr. Thierry Burdeaux, Mr. Nicolas Derrier from STM (France); Dr. Geoffrey Coram from ADI (USA), Dr. Colin McAndrew from Freescale (USA), Mr. Mu-Liang Xu from Nortel (Canada); Prof. Jean-Claude Perraud from CAEN (France), Mrs. Jobymol Jacob (now at IIT Madras). Special thanks go here to Mr. Anindya Mukherjee from CEDIC at TUD (Germany), who took over from the second author (AC) in 2007 the HICUM user support and implementation.

We are indebted to the German Research Foundation (DFG), the German Ministry of Education and Research (BMBF), the German Ministry of Economic Affairs (BMWi), and the European Community. Their *long-term research* funding has been a crucial complement to the usually short-term application oriented development work funded by industry. We just have one comment: considering the ever decreasing budgets, project durations, and degrees of freedom on one hand and the increasing shift of administrative and reporting burden to the research partners on the other hand, we would wish for a reversal of this trend sooner rather than later in order to enable the pursuit of more innovative approaches and to boast the efficiency of research and development work.

Finally, we would like to thank Mr. Dider Celi (STM), Mr. Andreas Pawlak, Mr. Steffen Lehmann, Mr. Anindya Mukherjee (all with CEDIC), Mr. Zoltan Huszka (AMS), Dr. Plamen Kolev (PKMS), and Mr. Khamesh Kumar (IIT Madras) for proofreading some of the chapters of this book and for providing valuable feedback. MS would like to thank Prof. Peter Asbeck and Prof. Larry Larsen, both of UCSD, for facilitating the extended stays during which most chapters of this book were written. Last but not least, we are indebted to Prof. Chih-Tang Sah for giving us the opportunity

to publish this book with World Scientific, and to Ms. Yubing Zhai for her continuous help with all publication related aspects.

**Michael Schröter**

**Anjan Chakravorty**

30th June, 2009

# Table of Contents

<b>Foreword</b>	vii
<b>Biographies</b>	1
<b>Preface</b>	3
<b>Chapter 1: <i>Introduction</i></b>	23
References	30
<b>Chapter 2: <i>Device Modeling Overview</i></b>	37
2.1 A brief history on bipolar transistor technology	39
2.2 Semiconductor equations	49
2.3 Compact modeling	55
2.4 Charge definitions in bipolar transistors	59
2.4.1 Regional analysis of bipolar transistors	61
2.4.2 Relation between charge storage components and measurement	71
References	73
<b>Chapter 3: <i>Theory of Homojunction Bipolar Transistors</i></b>	77
3.1 Operation principle	78
3.1.1 "High" collector voltage	79
3.1.2 "Low" collector voltage	83
3.1.3 Summary	86
3.2 Electric field calculations	87
3.2.1 Fully depleted collector region - high voltage case	88
3.2.2 Partially depleted collector: low voltage case	89
3.3 Static operation and characteristics	95
3.3.1 Transfer current	95
3.3.1.1 Low injection	97

3.3.1.2 High injection	103
3.3.1.3 Medium injection	106
3.3.2 Back injection related current components	108
3.3.3 Currents related to space-charge region recombination	114
3.3.4 Avalanche current from base-collector breakdown	117
3.4 Depletion (junction) charges and capacitances	123
3.4.1 Classical approach	123
3.4.2 High forward bias operation	127
3.4.3 Punch-through case	129
3.5 Mobile charge and storage times	131
3.5.1 Low injection	132
3.5.1.1 Neutral base region	133
3.5.1.2 Base-collector region	138
3.5.1.3 Neutral emitter	144
3.5.1.4 Base-emitter space charge region	148
3.5.1.5 Summary of the results for low current densities	152
3.5.2 Medium current densities	152
3.5.2.1 Neutral base region	153
3.5.2.2 Collector region	154
A. Low collector voltage	155
B. High collector voltage	156
C. Critical current	156
3.5.2.3 BC space-charge region	157
3.5.2.4 Neutral emitter region	158
3.5.2.5 BE space-charge region	161
3.5.2.6 Summary	161
3.5.3 High current densities	161
3.5.3.1 Collector region	162
A. Low collector voltage	162
B. High collector voltage	163
C. Charge and storage time	165
3.5.3.2 Neutral base region	169
3.5.3.3 BC space-charge region	171
3.5.3.4 Neutral emitter region	173
3.5.3.5 BE space-charge region	173
3.5.3.6 Summary	174

References	174
<b>Chapter 4: Advanced Theory</b>	179
4.1 HBT operation principle	180
4.1.1 The SiGe heterojunction in a neutral region	180
4.1.2 Heterojunction diode	184
4.1.3 SiGe HBT	187
4.1.3.1 High-voltage case	191
4.1.3.2 Low-voltage case	193
4.1.3.3 The base-collector barrier	195
4.2 Electric field in the collector region	203
4.3 Static operation and characteristics	209
4.3.1 Transfer current from integral charge-control relations	210
4.3.1.1 The Moll-Ross relation	214
4.3.1.2 The integral charge-control relation	215
4.3.1.3 Generalized integral charge-control relation	221
A. Hole charge partitioning	227
B. Depletion charge	228
C. Minority charge	232
D. Compact formulation	234
4.3.2 Base current components related to recombination	237
4.3.3 Non-local base-collector avalanche current	240
4.3.4 Tunnelling	247
4.4 Charge storage	249
4.4.1 Depletion charges	249
4.4.1.1 Influence of the material composition	249
4.4.1.2 Current dependence of base-collector depletion charge	253
4.4.2 Mobile charge	258
4.4.2.1 Neutral base region	259
4.4.2.2 Neutral collector	262
4.4.2.3 Base-collector depletion region	266
4.4.2.4 Neutral emitter	266
4.4.2.5 Base-emitter depletion region	269
4.5 Vertical non-quasi-static effects	270
4.5.1 Introduction	270
4.5.2 Time-domain description	272

4.5.3 Frequency domain solution	283
4.5.4 Discussion and relation to compact modeling	290
References	296
<b>Chapter 5: Geometry (Layout) Scaling</b>	<b>301</b>
5.1 Introduction	302
5.2 Internal base resistance	305
5.2.1 Static operation	306
5.2.1.1 Current crowding effect	306
5.2.1.2 Conductivity modulation	312
5.2.2 Small-signal operation	317
5.2.3 Large-signal transient operation	324
5.3 Emitter junction perimeter effects	332
5.3.1 Standard scaling of currents and charges	333
5.3.2 Transfer current	335
5.3.3 Base current components	354
5.3.4 Charge storage	357
5.3.4.1 Depletion charges	358
5.3.4.2 Minority charge	359
5.3.4.3 Base resistance under the BE spacer	361
5.4 Three-dimensional and parasitic effects	363
5.4.1 Standard-scaling for a 3D structure	364
5.4.2 Minority charge	365
5.4.3 Base resistance	372
5.4.3.1 Internal base resistance for the 3D case	377
A. Static operation	377
B. Small-signal frequency dependent operation	380
C. Large-signal transient operation	382
5.4.3.2 Total base resistance for the 3D case	382
A. Double base contact in parallel to the emitter	383
B. Single base contact in parallel to the emitter	388
C. Base contacts perpendicular to the emitter	394
5.4.3.3 Summary	397
5.4.4 Avalanche pinch-in effect	398
5.4.5 Base-collector region related capacitances	402
5.4.6 Base-emitter isolation capacitance	404

5.4.7 External collector resistance	405
5.4.8 Emitter resistance	409
5.4.9 Collector-substrate capacitance	411
5.4.10 Parasitic substrate transistor	413
5.4.11 Intra-device substrate coupling	416
5.5 Non-standard-scaling	419
References	424
<b>Chapter 6: <i>Temperature Effects</i></b>	431
6.1 Overview	432
6.2 Physical effects	432
6.3 Temperature dependence of transistor characteristics	438
6.3.1 Charge storage	438
6.3.1.1 Depletion charge	438
6.3.1.2 Mobile and minority charge	442
6.3.1.3 Zero-bias hole charge	444
6.3.2 Current components	447
6.3.2.1 Transfer current	447
6.3.2.2 Back injection related base current components	449
6.3.2.3 Recombination currents	450
6.3.2.4 Base-collector avalanche current	451
6.3.2.5 Temperature dependence of current gain	451
6.3.2.6 Tunneling current	453
6.3.3 Series resistances	453
6.3.4 Noise	454
6.4 Self-heating effects	454
6.4.1 Intra-device electro-thermal coupling	456
6.4.2 Inter-device electro-thermal coupling	461
References	465
<b>Chapter 7: <i>Compact Noise Modeling</i></b>	471
7.1 Basic noise mechanisms	472
7.1.1 Fundamentals	472
7.1.2 Thermal noise	477
7.1.3 Shot noise	480
7.1.4 Recombination and generation noise	481

7.1.5 Flicker noise	483
7.2 Intrinsic transistor	485
7.2.1 Neutral base region	485
7.2.2 Base-collector space-charge region	493
7.2.3 Neutral emitter	494
7.2.4 Impact ionization noise	494
7.3 Multi-dimensional effects	495
7.3.1 Internal transistor	495
7.3.2 External (access) regions	496
7.3.3 Flicker and 1/f noise	497
References	500
<b>Chapter 8: HICUM Level2</b>	<b>503</b>
8.1 Internal transistor model	504
8.1.1 Depletion charges and capacitances	506
8.1.1.1 Internal base-emitter junction	506
8.1.1.2 Internal base-collector junction	507
8.1.2 Minority charge and transit times	511
8.1.2.1 Forward minority charge component	512
8.1.2.2 Inverse minority charge component	519
8.1.3 Quasi-static transfer current	519
8.1.4 Non-quasi-static effects	523
8.1.5 Quasi-static base current components	525
8.1.6 Collector-base breakdown	527
8.1.7 Emitter-base tunneling	528
8.1.8 Internal base resistance	529
8.2 Complete transistor	533
8.2.1 Mobile charge including collector current spreading	534
8.2.2 Emitter perimeter region	538
8.2.3 External base-collector region	540
8.2.4 External series resistances	541
8.2.5 Collector-substrate junction	542
8.2.6 Parasitic substrate transistor	542
8.2.7 Substrate network	543
8.3 Temperature dependence	544
8.3.1 Temperature dependent bandgap voltage	545



8.3.2 Depletion charges and capacitances	545
8.3.3 Transfer current	547
8.3.4 Transit time and minority charge	548
8.3.5 Junction current components	550
8.3.6 Series resistances	553
8.3.7 Parasitic substrate transistor	554
8.4 Self-Heating model	554
8.5 Small-signal model	557
8.6 Noise model	559
8.7 Model parameter list	564
References	576
<b>Chapter 9: Parameter Determination for HICUM/L2</b>	<b>579</b>
9.1 Introduction	580
9.2 Wafer selection	582
9.3 Relevant transistor dimensions	583
9.4 Measurements	585
9.4.1 IV measurements and data	585
9.4.2 CV measurements and data	585
9.4.3 S-parameter measurements and data	586
9.4.4 Measurement conditions	587
9.4.5 Sequence of measurements	590
9.5 Extraction flow	590
9.6 Step-by-step extraction procedure	592
9.6.1 BC depletion and isolation capacitance	596
9.6.2 BE depletion and isolation capacitance	598
9.6.3 CS depletion capacitance	599
9.6.4 Internal base (sheet) resistance	600
9.6.5 Components of external base resistance	601
9.6.6 Emitter resistance	602
9.6.7 Components of external collector resistance	603
9.6.8 Collector current at low bias	604
9.6.9 Current across BE junction at low bias	606
9.6.10 Current across the BC junction (at low bias)	607
9.6.11 Thermal resistance	608
9.6.12 Forward transit time	609

9.6.13 Collector current at high injection	611
9.6.14 Base-collector Breakdown	613
9.6.15 High-Frequency effects	614
9.6.15.1 Vertical non-quasi-static effects	614
9.6.15.2 Partitioning of the external BC capacitance	615
9.6.16 Intra-device substrate coupling	616
9.6.17 High-frequency emitter current crowding	617
9.6.18 Parasitic substrate transistor elements	618
9.6.18.1 Transfer current	618
9.6.18.2 Charge storage time	619
9.6.19 Temperature dependence	619
9.6.19.1 Bandgap voltages related parameters	620
9.6.19.2 Series resistances	621
9.6.19.3 Transit time at low current densities	621
9.6.19.4 Collector-emitter saturation voltage	622
9.6.19.5 BC breakdown	622
9.7 Summary and discussions	624
References	626
 <b>Chapter 10: Model Hierarchy</b>	 631
10.1 Introduction	632
10.2 HICUM/L0: a simplified model??	634
10.2.1 Large-signal model formulation	635
10.2.1.1 Equivalent circuit	636
10.2.1.2 Depletion charges and capacitances	637
A. Base-emitter junction	637
B. Base-collector junction	638
C. Collector-substrate junction	640
10.2.1.3 Minority charge	640
10.2.1.4 Quasi-static transfer current	642
10.2.1.5 Static base current components	652
10.2.1.6 Avalanche current	653
10.2.1.7 Parasitic substrate transistor	654
10.2.1.8 Series resistances	654
10.2.1.9 External (parasitic) capacitances	657
10.2.1.10 Self-heating	657

10.2.2 Temperature dependence	658
10.2.3 Small-Signal Operation	661
10.2.4 Noise Model	662
10.2.5 Parameter List	663
10.2.6 Parameter Extraction	669
10.2.6.1 Parameter determination from HICUM/L2	669
10.2.6.2 Parameter extraction from experimental data	670
10.3 HICUM/L4: a distributed mode	671
References	676
<b>Chapter 11: <i>Application Examples</i></b>	679
11.1 Geometry scaling approach	680
11.2 Examples for comparisons to measured device characteristics	685
11.2.1 HICUM/L2	685
11.2.2 HICUM/L0	690
11.2.3 HICUM/L4	691
11.2.4 Statistical and predictive modeling	692
11.3 Model deployment for circuit design	700
References	706
<b>Chapter 12: <i>Future Trends</i></b>	715
12.1 Technology challenges	716
12.2 Future applications	721
12.3 Technology trends	722
12.4 Physical limitations and device modeling	728
References	730
<b>Index</b>	735

This page is intentionally left blank

# List of Often Used Acronyms and Symbols

## 1.1 List of often used acronyms

AC, DC	alternating, direct current
BE, BC	base-emitter, base-collector
BGN	bandgap narrowing
BiCMOS	bipolar and complementary metal oxide semiconductor
BJT	bipolar junction transistor
BTB	band-to-band tunneling
BTE	Boltzmann transport equation
CE, CS	collector-emitter, collector-substrate
CED	conventional emitter doping
CML	current-mode logic
CMOS	complementary metal oxide semiconductor
DD, DDF	drift diffusion, DD formulation
DTI	deep trench isolation
EC	equivalent circuit
FN	flicker noise
Ge	germanium
GICCR	generalized integral charge control relation
HBT	heterojunction bipolar transistor
HD, HDF	hydro dynamic, HD formulation
HF	high-frequency
HICUM/L2	high current model/ level 2
IC	integrated circuit
ICCR	integral charge control relation
LEC	low-emitter concentration
LNA	low noise amplifier

MC	Monte Carlo
MOSFET	metal oxide semiconductor field effect transistor
NQS	non-quasi-static
NR	neutral region
PA	power amplifier
PBSA	process based scalable approach
PCM	process control monitor
PCB	partitioned charge based
PDK	process design kit
QS	quasi-static
QFP	quasi-Fermi potential
RA	regional approach
r.m.s.	root mean square
SCR	space charge region
SGPM	SPICE Gummel Poon model
Si	silicon
SIC	selectively implanted collector
SiGe	silicon germanium
TC	temperature coefficient
TICCR	transient integral charge control relation
TRADICA	transistor dimensioning and calculation program
VA	Verilog-AMS
1D, 2D, 3D	one-, two-, three-dimensional

## 1.2 List of often used symbols

<sup>1</sup>Small case letters for currents and voltages express time dependent variables, large case letters bias dependence, and underscored large-case letters complex (frequency domain) variables.

$A_{E0}, A_E$	emitter area, effective electrical emitter area
$b_{E0}, b_E$	emitter width, effective electrical emitter width
$BV_{CEO}$	collector emitter breakdown voltage with open circuited base

- 
1. Variable designations correspond (as much as possible) to those used in HICUM related publications.

$C_{jE}, C_{jC}$	base-emitter and base-collector junction capacitances
$\delta_C$	collector current spreading angle
$E_x (\bar{E}_x)$	semiconductor region dependent (average) electrostatic field
$E_g (V_g)$	Bandgap energy (voltage)
$f_{cs}$	collector current spreading factor
$f_T, f_{max}$	cut-off frequency, maximum oscillation frequency
$g, C, \tau, \omega$	conductance, capacitance, transit time, and frequency in radian
$h$	GICCR weight factors
$I_C, I_T, I_B$	bias dependent collector, transport, and base current
$J_C (J_{CK})$	(critical) collector current density
$J_n, J_p$	electron and hole current density
$l_{E0}, l_E$	emitter length, effective electrical emitter length
$M$	multiplication factor
$\gamma_C$	ratio of periphery to area specific collector current
$\mu_n, \mu_p (\bar{\mu}_n, \bar{\mu}_p)$	electron, hole (average) mobility
$n, p$	electron, hole concentration
$n_i$	intrinsic concentration of electron
$N_E, N_B, N_{Ci}$	emitter, base, and epi-collector doping concentration
$NF_{min}$	minimum noise figure
$Q_{jE}, Q_{jC}$	base-emitter, base-collector space charge region charges
$Q_p, Q_m$	total hole and minority charges
$r_{SBi}$	internal base sheet resistance
$T (T_0)$	ambient (nominal or room) temperature
$V_{B'E'}, (V_{BE})$	potential difference between internal base and internal emitter (base and emitter) nodes
$V_{B'C'}, (V_{BC})$	potential difference between internal base and internal collector (base and collector) nodes
$V_{C'E'}, (V_{CE})$	potential difference between internal collector and internal emitter (collector and emitter) nodes
$V_{ci}$	voltage dropped in the epi-collector region
$V_{CEs}$	collector-emitter saturation voltage
$V_T$	thermal voltage
$w_E, w_B$	neutral region width for emitter and base
$w_{BE}, w_{BC}$	width for base-emitter and base collector

	space charge region
$w_{Ci}$	(effective) collector width under emitter
$w_i$	width of collector injection zone
$\varphi_n, \varphi_p$	quasi-Fermi potential for electron and hole
$\psi_n$	electro-static potential for electron
$x, y, z$	three dimensions for device
$x_{jE}, x_{jC}$	metallurgical junction point for base-emitter and base-collector pn-junction
$n, i, v, Q$	time-dependent large-signal variables: electron conc., current, voltage, and charge
$\underline{n}, \underline{I}, \underline{V}, \underline{Q}$	frequency-dependent small-signal variables: electron conc., current, voltage, and charge



# Chapter 1

## Introduction

Although CMOS has been the mainstream technology over more than the past two decades, even today silicon-based bipolar transistors are still leading their field-effect counterparts in terms of operating speed for both devices and circuits<sup>1</sup>. This speed advantage seemed to fade in the late eighties, when stand-alone bipolar technology lost ground to CMOS due to the use of depreciated and old fabrication tools, but became obvious again by the mid nineties, when the silicon-germanium (SiGe) heterojunction bipolar transistor (HBT) was developed and subsequently also integrated into a CMOS baseline process. Since then the resulting BiCMOS technology has gained tremendous success as can be observed from the number of commercial fabrication facilities, which has grown within the past decade from one in the mid nineties to over 30 (e.g. [1]). In addition, there are numerous (> 20) companies worldwide with facilities for fabricating HBTs in III/V materials such as GaAs, InP, and GaN. Finally, both HBTs and silicon homojunction bipolar transistors (BJTs) still capture a quite sizeable market for discrete linear analog applications, while silicon-carbide HBTs have positioned themselves quite well for high-voltage and high-temperature operation in the automotive and high-power market.

In fact, measured just by volume, the discrete market still has quite a substantial share in shipped items. Its 2008 revenue is about 7% of the \$273 billion semiconductor industry, with an average annual growth rate

---

1. Assuming both transistor types are realized in the same material system and lithography node; i.e. the same critical dimensions. For more details and data regarding this comparison see chapter on “Future Trends”.

of 5%. Overall, the analog integrated circuits (ICs) and discrete market together constitute 21 percent. Interestingly, over the next 5 years the analog IC market is expected to have an average annual growth rate of 12% and to outperform the expected growth rates for both the digital market (6%) and the global market (7%) [2].

Due to their superior analog high-frequency (HF) characteristics and performance (e.g. [3]), bipolar transistors have maintained a strong foothold in corresponding application areas such as communications, automotive, measurement equipment, and power control. In fact, the linear and discrete market is still dominated by bipolar technology. The performance and functionality of *integrated* circuits and systems have benefitted significantly from the capability of BiCMOS technology, which often offers an early and cost efficient introduction of HF/high-speed products. It is worth noting that the use of Ge is beneficial to the performance of SiGe HBTs to a much higher extent than in strained SiGe MOSFETs. While in both transistor types the strain caused by introducing Ge increases the mobility, the vertical layer structure in HBTs also allows to create aiding fields and trade off, e.g., current gain versus speed through energy gap (bandgap<sup>2</sup>) engineering. Hence, it does not come at a surprise that SiGe BiCMOS technology has found numerous applications (cf. [1] and the chapter on future trends in this book). As a consequence, SiGe *HBTs* are the main topic of this book although their theoretical treatment of course is built on the generic fundamentals applying also to BJTs.

Circuit design and optimization today are heavily based on simulation. In other words, a circuit designer generally observes the features of a semiconductor process technology through a so-called process design kit (PDK). In order to reflect reality as much as possible and, hence, to minimize design iterations and cost, the PDK must contain very accurate models of the devices fabricated in the underlying process. On the other hand such models need to be sufficiently simple for circuit simulation in terms of computational effort, i.e. simulation time, if analog HF circuits with hundreds or up to many thousands of different devices have to be simulat-

---

2. Although the physically correct expression here is energy gap, the designation bandgap will be used in the book since it has been in use for many years in the existing (engineering) literature.

ed in time domain. Device models that are suitable for such tasks are generally known under the term *compact* models (for a more detailed definition and discussion see sec. 2.3).

Optimizing analog HF circuit performance with respect to given specifications is generally achieved by *device sizing* (e.g. [4-7]); i.e. by varying the transistor *configuration* consisting of emitter dimensions and contact arrangement. For a given process, this is generally the only way to achieve a proper trade-off between electrical transistor characteristics or parameters that often countertrend each other. Figure 1.1 visualizes the difficulties in the optimization task for two different types of circuits, a low noise HF amplifier and a CML circuit. In a low-noise amplifier (LNA), the lowest noise is usually obtained close to the minimum of the transistor's minimum noise figure  $NF_{min}$ . However,  $NF_{min}$  occurs at a much lower collector current density than the peak of the transit frequency  $f_T$  or the maximum oscillation frequency  $f_{max}$ , which indicate the maximum speed of the *device itself*. Hence, in an LNA minimum noise and maximum operating frequency (or bandwidth) can generally not be achieved simultaneously. Instead one has to trade-off one against the other by varying the collector current density at a given bias current  $I_C$  that is usually defined by power dissipation or other constraints. This trade-off is performed by varying the emitter length [4,6,7] and, thus, the associated emitter area  $A_E$ .

In "digital" circuits such as frequency dividers and multiplexers, the goal is often to minimize gate delay  $\tau_d$ . Depending on the load (or fan-out)  $\tau_d$  usually assumes its minimum beyond the peak of  $f_T$  or  $f_{max}$  (see Fig. 1.1). Since minimum  $\tau_d$  requires (current switch) transistors to be operated at high current densities additional jitter might be generated [8]. Therefore, minimizing both  $\tau_d$  and jitter requires a careful selection of the proper transistor *configuration* [4, 9]. Furthermore, different transistor sizes are generally also necessary in bias circuits. As a consequence of the above discussion, compact models need to be geometry scalable, which means that *every* element in an equivalent circuit (EC) has to be described as a function of the transistor *configuration*. In addition, design houses demand statistical modeling (and design) capability from foundries. Going even further, competition sometimes forces foundries to release compact models that *predict* the targeted performance with sufficient accuracy well

before process qualification. These goals can only be accomplished by using *physics-based* compact models.

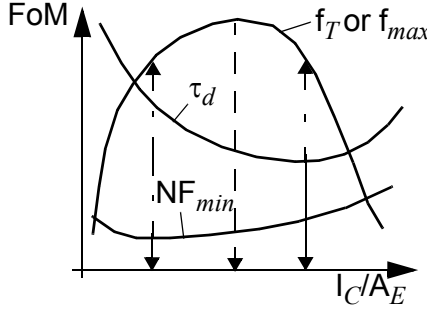


Fig. 1.1: Circuit optimization through trade-off between countertrend characteristics indicated here by important figures of merit vs. collector current density.

The advanced compact model HICUM meets the before mentioned challenges. Its development started in the early eighties (e.g. [10-12]), when it was realized that the charge storage description and scalability of the SPICE Gummel-Poon model (SGPM) were inaccurate and inadequate for designing high-speed fiber-optic circuit components. For such applications operation at high current densities is very important. This bias region, in which the SGPM is particularly inaccurate, was the focus of the initial model development. So, the name HICUM was derived from *High-CU*rent *Model*.

Since geometry scaling was not (even until today) part of a compact *bipolar transistor model implementation* in circuit simulators, the corresponding equations were implemented in a separate tool named TRADICA [13] in order to facilitate transistor sizing and to generate scalable model parameters for designing high-speed circuits (e.g. [4]). Early versions of HICUM were integrated in various SPICE2 codes, and the first comprehensive experimental verification of the large-signal dynamic behavior were performed in [14-16].

In the early nineties an accurate small-signal formulation was developed and experimentally verified [17], immediately followed by first extensions for including effects in SiGe HBTs [18] and three-dimensional geometry scaling of collector current flow [19, 20]. Work until the end of the nineties comprised further extensions and experimental verifications especially re-

lated to wireless (as opposed to fiber-optic) applications (e.g. [21-28]). Demand for a more widespread simulator implementation caused the effort to shift for a while to the area of numerical stability, efficient coding and simulator testing. Once the model was available in certain commercial simulators, the development work in this decade has focused again more on physics-based extensions and improvements (e.g. [29-43]), parameter extraction issues (e.g. [33-39]) and experimental verifications including noise (e.g. [40-43]) as well as electrothermal modeling (e.g. [44-49]) and statistical modeling (e.g. [50-52]). Moreover, experience with III/V HBTs has been gained in e.g. [53-59].

During the mid nineties, first implementations into commercial circuit simulators took place but, due to cumbersome interfaces, were limited to to in-house company codes and mainstream programs such as ELDO and SPECTRE. In order to overcome these obstacles one of the authors of this book asked EDA companies already during the early nineties to automate the implementation process and provide model “compilers”. However, the latter took over a decade to become publicly available. Also, in the late nineties the Compact Model Council (CMC) was formed with the goal to simplify model implementation and to standardize interfaces so as to obtain the *same implementation and results across all simulators*.

Early on in this decade the CMC evaluated compact bipolar transistor models for several years on six different process technologies. In 2003, HICUM was selected as one of two standard models. Since at that time there was still no unified simulator interface, the purpose of the selection of models as standard was to allow EDA vendors to focus on the implementation and maintenance of *very few* models in order to meet the original goals of the CMC (i.e. same model results across simulators). Only when Verilog-AMS (VA) had advanced far enough model compilers started to become available [60, 61]. These compilers take high-level (VA) code as input, translate it to C code, adapt it to the still different simulator interfaces, and then compile the C code. This allows any compact model to be linked to any of the (supported) circuit simulators and has significantly reduced the implementation burden for model developers.

The selection of HICUM as a standard expectably has also led to an increase of users in the foundry and design community. By now, there exist experimental verifications of all sorts for many different bipolar process

technologies, and HICUM has become available in many PDKs (e.g. [62]). As a result, the model has been employed in production designs fabricated at leading SiGe foundries around the world for quite a while now (see chapter 11 on applications).

So far, the name HICUM has been used synonymously for HICUM/Level2 (L2) [63], which is the more precise designation for the original model mentioned so far. Version 2.0 was the first that was widely implemented as standard model in commercial simulators while version 2.2 was the first compact model to be implemented entirely via a VA description *and* model compilers.

The complexity of compact models can vary widely depending on the application focus, but generally has increased significantly as a result of an ever increasing complexity of device designs and associated number of physical effects. Therefore, existing standard models for bipolar transistors such as HICUM/L2 contain a representation of all known relevant physical effects, albeit in simplified form. Hence, the resulting model (i.e. equivalent circuit and equations) appears to be fairly complicated to circuit designers and other users<sup>3</sup>, and can also lead to long simulation times for *large* circuits. This has generated a demand for a simplified model. On the other hand, transistor structures employed in driver circuits and power amplifiers (PAs) tend to become quite large and often are realized as multi-finger transistors or are arranged as arrays of fixed-finger transistors (i.e. cells). The resulting distributed electrical and thermal effects cannot always be captured sufficiently accurately by lumped models, even those like HICUM/L2. Meeting these very different design needs makes a compact model *hierarchy* very attractive (cf. chapter 10). This way, circuit designers can select the model level that for a given task is the most suitable with respect to simulation speed and accuracy.

Minimizing overall product development cost requires PDKs with accurate compact models, especially for HF circuit design (e.g. [4-9]). The in fact increasing importance of the demand for accurate models can be easily understood from Fig. 1.2. Already in 90nm CMOS and BiCMOS processes the overall mask cost exceed US\$ 1Mio, and the existing trend points to a rapid increase in cost for more advanced processes. To put this in per-

---

3. This is true also for MOS transistor models.

spective, one can say that saving overall just one complete design iteration and its associated mask cost by providing accurate models already pays for the annual expense of a typical compact modeling group in industry! Therefore, adequate investment in the development of accurate compact models does indeed make a lot of sense today also from a business perspective.

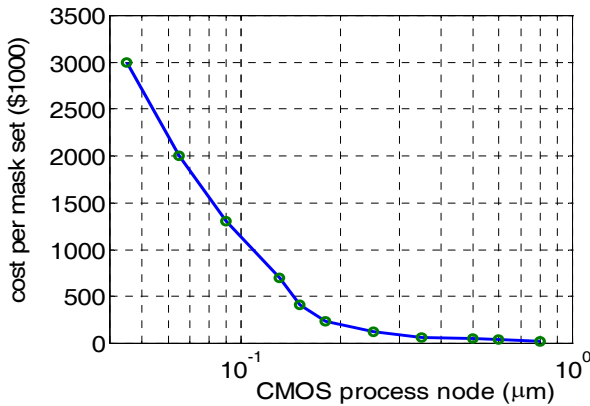


Fig. 1.2: Typical fabrication cost development, measured in terms of total cost per mask set, as function of the CMOS process node, measured by critical dimension.

It is interesting to note though that those model users, namely the circuit design and EDA industry, which benefit directly from good compact models, provide the least amount of support compared to wafer manufacturers. However, even the latter still focus almost entirely on the *directly design related* demand by investing just in parameter extraction for a production process, but struggle with investments in the *development of model improvements* for new effects and with establishing the modeling related infrastructure changes early enough for an upcoming new process. There may be some hope for improvement of this situation as the understanding of the consequences of Fig. 1.2 apparently have led to increasing support of MOS compact model development.

## 1.1 References

- [1] J. Cressler (ed.), "Silicon heterostructure Handbook", CRC Press NY, 2005.
- [2] [www.electronics.ca/reports/ic/global\\_semiconductors.html](http://www.electronics.ca/reports/ic/global_semiconductors.html)
- [3] D. Harame, "RF Device Technologies", IEDM Short Course, 2002.
- [4] H.-M. Rein, M. Möller, "Design considerations for very-high-speed Si-bipolar ICs operating up to 50Gb/s", IEEE J. Solid-State Circuits, Vol.31, pp. 1076-1090, 1996.
- [5] T. Massier, H. Graeb, U. Schlichtmann, "The sizing rules method for CMOS and bipolar analog integrated circuit synthesis", IEEE Trans. Comp.-Aided Des. Integr. Circ. Syst., Vol. 27, pp. 2209-2222, 2008.
- [6] S. Voinigescu, M. Maliepaard, M. Schröter, P. Schvan and D. Harame, "A scalable high-frequency noise model for bipolar transistors and its application to optimal transistor sizing for low-noise amplifier design", IEEE J. Solid-State Circuits, Vol. 32, no. 9, pp. 1430-1439, 1997.
- [7] J. Long, "SiGe Radio frequency ICs for low-power portable communication", Proc. IEEE, Vol. 93, pp. 1598-1623, 2005.
- [8] M. Schröter, H.-M. Rein, "Investigation of very fast and high-current transients in digital bipolar circuits by using a new compact model and a device simulator", IEEE J. Solid-State Circuits, Vol. 30, no. 5, pp. 551-562, 1995.
- [9] H. Li, H.-M. Rein, "A 77/100 GHz VCO in SiGe technology", IEEE J. Solid-State Circuits, Vol. 39, pp. 1650-1658, 2004.
- [10] M. Schröter and H.-M. Rein, "Two-dimensional modelling of high-speed bipolar transistors at high current densities using the Integral Charge-Control Relation relation", Proc. ESSDERC '84; see also: Physica B129, North Holland Phys. Publ. Div., pp. 332-336, 1985.
- [11] H.-M. Rein, H. Stübing, M. Schröter, "Verification of the Integral Charge-Control Relation for high-speed bipolar transistors at high current densities", IEEE Trans. Electron Dev., Vol. 32, no. 6, pp. 1070-1076, 1985.
- [12] M. Schröter, H.-M. Rein, "A compact physical large-signal model for high-speed bipolar transistors including the high-current region", NTG meeting, Würzburg, Mai 1986.
- [13] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener, A. Koldehoff, "Physics- and process-based bipolar transistor modeling



- for integrated circuit design", IEEE Journal of Solid-State Circuits, Vol. 34, no. 8, pp. 1136-1149, 1999.
- [14] H.-M. Rein, M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors at high current densities - Part II: Two-dimensional model and experimental results", IEEE Trans. Electron Dev., Vol. 34, no. 8, pp. 1752-1761, 1987. se also:  
H. Stübing, H.-M. Rein, "A compact physical large-signal model for high-speed bipolar transistors at high current densities - Part I: One-dimensional model", IEEE Trans. Electron Dev., Vol. 34, no. 8, pp. 1741-1751, 1987.
- [15] M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors with special regard to high current densities and two-dimensional effects", Ph. D. thesis (in German), Ruhr-University Bochum, Bochum, Germany, 1988.
- [16] M. Schröter, H.-M. Rein, "Transit time of high-speed bipolar transistors in dependence on operating point, technological parameters, and temperature", Proc. IEEE Bipolar Circuits and Technology Meeting, Minneapolis, pp. 250-253, 1989.
- [17] A. Koldehoff, M. Schröter, H.-M. Rein, "A compact bipolar transistor model for very high-frequency applications with special regard to narrow stripes and high current densities", Solid-State Electronics, Vol. 36, pp. 1035-1048, 1993.
- [18] M. Schröter, M. Friedrich, H.-M. Rein, "A generalized Integral Charge-Control Relation and its application to compact models for silicon based HBT's", IEEE Trans. Electron Dev., Vol. 40, no. 11, pp. 2036-2046, 1993.
- [19] M. Schröter and D. J. Walkey, "Scaling considerations of bipolar transistors using 3D device simulation", Proc. SISDEP, pp. 210-213, Sept. 1995.
- [20] M. Schröter, D. J. Walkey, "Physical modeling of lateral scaling in bipolar transistors", IEEE J. Solid-State Circuits, Vol. 31, no. 10, pp. 1484-1491, 1996 and Vol. 33, p. 171, 1998.
- [21] M. Schröter, Z. Yan, T.-Y. Lee, W. Shi, "A compact tunneling current and collector breakdown model", Proc. IEEE Bipolar Circuits and Technology Meeting, Minneapolis, pp. 203-206, 1998.
- [22] M. Schröter, T.-Y. Lee, "A physics-based minority charge and transit time model for bipolar transistors", IEEE Trans. Electron Dev., vol. 46, no. 2, pp. 288-300, 1999.

- [23] M. Schröter, D.R. Pehlke and T.-Y. Lee, "Compact modeling of high-frequency distortion in bipolar transistors", Proc. ESSDERC, Leuven, pp. 476-479, 1999.
- [24] M. Schröter, D.R. Pehlke, T.-Y. Lee, "Compact modeling of high-frequency distortion in Si integrated bipolar transistors", IEEE Trans. on Electron Dev., Vol. 47, pp. 1529-1539, 2000.
- [25] H.Q. Tran, M. Schröter, D. J. Walkey, D. Marchesan and T. J. Smy, "Simultaneous extraction of thermal and emitter series resistances in bipolar transistors", Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting, Minneapolis, pp. 170-173, 1997.
- [26] X.Y. Chen, M. J. Deen, Z.X. Yan and M. Schröter, "Effects of emitter dimensions on low-frequency noise in double-polysilicon BJTs", Electronics Letters, Vol. 34, No. 2, pp. 219-220, 1998.
- [27] X. Y. Chen, M. J. Deen, A.D. van Rheeën, Z. X. Yan, and M. Schröter, "Low-frequency noise in npn and pnp double-polysilicon BJTs, effects of temperature and emitter dimensions", Proc. Int'l Electron Devices and Materials Symposium (IEDMS), 4 pages, 1998.
- [28] M. J. Deen, S. L. Rumyantsev and M. Schröter, "On the origin of  $1/f$  noise in polysilicon emitter bipolar transistors", J. Appl. Phys., Vol. 85, No. 2, pp. 1192-1195, 1999.
- [29] M. Schröter, H. Tran, "Modeling of charge and collector field in Si-based bipolar transistors", Proc. WCM, International NanoTech Meeting, Boston (MA), pp. 102-107, 2004.
- [30] M. Schröter and H. Tran, "2D/3D-GICCR for Si/SiGe bipolar transistors", acc. for publication at the Proc. Compact Modeling Workshop at the International NanoTech Meeting, Anaheim (CA), May 2005.
- [31] P. Sakalas, A. Chakravorty, J. Herricht, M. Schröter, "Compact Modeling of High Frequency Correlated Noise in HBTs", Proc. Bipolar Circuits and Technology Meeting (BCTM), Maastricht (Belgium), pp. 279-282, 2006.
- [32] P. Sakalas, M. Ramonas, M. Schröter, C. Jungemann, A. Shimukovich and W. Kraus, "Impact Ionization Noise in SiGe HBTs: comparison of device and compact modeling with experimental results", IEEE Trans. Electron Devices, Vol. 56, No. 2, pp. 328-336, 2009.
- [33] see various presentations at HICUM Workshops since 2000: [http://www.iee.et.tu-dresden.de/iee/eb/eb\\_homeee.html](http://www.iee.et.tu-dresden.de/iee/eb/eb_homeee.html)

- [34] D. Berger, D. Celi, M. Schröter, M. Malorny, T. Zimmer, and B. Ardouin, "HICUM parameter extraction flow for a single transistor geometry", IEEE BCTM, Monterey (CA), pp. 116-119, 2002.
- [35] M. Malorny, M. Schröter, D. Celi, D. Berger, "An improved method for determining the transit time of Si/SiGe bipolar transistors", Proc. BCTM, pp. 229-232, 2003.
- [36] M. Schröter, S. Lehmann, D. Celi, "Non-standard geometry scaling effects in high-frequency SiGe bipolar transistors", Proc. WCM, International NanoTech Meeting, Santa Clara, pp. 603-608, 2007.
- [37] M. Schröter, S. Lehmann, "The rectangular bipolar transistor tetra-rode structure and its application", Proc. ICMTS, Tokyo, Japan, pp. 206-209, 2007.
- [38] D. Berger, "Evaluation and verification of a transistor model for high-frequencies", Ph.D. thesis, Bordeaux University, 2004.
- [39] XMOD Technologies, see [www.xmodtech.com](http://www.xmodtech.com).
- [40] P. Sakalas, M. Schröter, P. Zampardi, M. Racanelli, "Modeling of SiGe power HBT intermodulation distortion using HICUM", Proc. ESSDERC, Lisboa, pp. 311-314, 2003.
- [41] P. Sakalas, M. Schröter, R. Scholz, H. Jiang, M. Racanelli, "Analysis of microwave noise sources in 150GHz SiGe HBTs", RFIC Symp., Tech. Dig., pp. 291-294, 2004.
- [42] P. Sakalas, M. Schröter, P. Zampardi, M. Racanelli, "Microwave noise in III-V and SiGe-based HBTs: comparison, trends, numbers", (inv. paper), 18th Int. Conf. on Noise and Fluctuations, Canary Islands (Spain), pp. 151-163, 2004.
- [43] P. Sakalas, A. Chakravorty, M. Schröter, M. Ramonas, J. Herricht, A. Shimukovitch, C. Jungemann, "Modeling of High Frequency Noise in SiGe HBTs", Proc. SISPAD, Monterey, CA, pp. 271-274, 2006.
- [44] D. J. Walkey, T. J. Smy, H. Tran, D. Marchesan, M. Schröter, "Prediction of thermal resistance in trench isolated bipolar device structures", Proc. IEEE Bipolar and BiCMOS Circuits and Technology Meeting, Minneapolis 12.3, pp. 207-208, 1998.
- [45] D. J. Walkey, T. J. Smy, D. Marchesan, H. Tran, C. Reimer, T.C. Kleckner, M.K. Jackson, M. Schröter and J.R. Long, "Extraction and modeling of thermal behavior in trench isolated bipolar structures", IEEE Bipolar and BiCMOS Circuits and Technology Meeting, Minneapolis, pp. 97-100, 1999.

- [46] D. J. Walkey, T. J. Smy, D. Marchesan, H. Tran, M. Schröter, “A scalable thermal model for trench-isolated bipolar devices”, *Solid-State Electronics*, Vol. 44, No. 8, pp. 1373-1380, 2000.
- [47] D. J. Walkey, T. J. Smy, C. Reimer, M. Schröter, H. Tran, D. Marchesan, M. Jackson, and T. Kleckner, “Modeling thermal resistance in trench-isolated bipolar technologies including trench heat flow”, *Solid-State Electronics*, Vol. 46, No. 1, pp. 7-17, 2002.
- [48] Y. Zimmermann, “Modeling of spatially distributed and sizing effects in high-performance bipolar transistors”, Master (diploma) Thesis, Chair for Electron Devices and Integrated Circuits, TU Dresden, 2004.
- [49] H. Beckrich, “Contribution to the characterization and modeling of integrated power bipolar transistors in a BiCMOS technology”, Ph.D. thesis, Bordeaux University, 2006.
- [50] M Schröter, H. Wittkopf, W. Kraus, “Statistical modeling of bipolar transistors”, *Proc. Bipolar Circuits and Technology Meeting (BCTM)*, Santa Barbara (CA), pp. 54-61, 2005.
- [51] H.-J. Wassener, H.-J. Strobel, D. Eichel, W. Schneider, M. Schröter, H. Wittkopf, “Statistical Simulation: Use of a TRADICA based Statistical Design Kit in the Cadence Environment”, *GMM Fachbericht “Analog ’06”*, Vol. 196, VDE Verlag, Dresden, pp. 233-236, 2006.
- [52] W. Schneider, M. Schröter, W. Kraus, H. Wittkopf, “Statistical simulation of high-frequency bipolar circuits”, *Proc. DATE07*, Nice, pp. 1397-1402, 2007.
- [53] P. Sakalas, M. Schröter, P. Zampardi, H. Zirath, R. Welser, “Microwave Noise Sources in AlGaAs/GaAs HBTs”, *IEEE International Microwave Symposium*, Seattle (WA), pp. 2117-2120, 2002.
- [54] P. Sakalas, M. Schröter, P. Zampardi, H. Zirath, “Shot noise reduction in AlGaAs/GaAs and InP/GaAs based HBTs”, *Proc. SPIE Fluctuations and Noise*, Santa Fe (NM), eds. J. Deen et al., pp. 387-396, 2003.
- [55] P. Sakalas and M. Schröter, “Analysis of the Microwave Noise Sources in HEMTs, MOSFETs and HBTs”, *17th Int. Conf. on Noise and Fluctuations*, Prague, pp. 247-250, 2003.
- [56] Y. Zimmermann, “Modeling and scaling of III-V HBTs using HICUM and TRADICA”, Student Project Report, Skyworks Inc., Newbury Park, CA, 2003, and Chair for Electron Devices and Integrated Circuits, TU Dresden.

- [57] Y. Zimmermann, M. Schröter, P. Zampardi, A. Klimashov, G. Tkachenko, “III-V HBT Modeling, scaling and parameter extraction using TRADICA and HICUM”, IEEE Workshop on Power Amplifiers, San Diego, CA, 2003.
- [58] P. Sakalas, M. Schröter, P. Zampardi, M. Racanelli, “Microwave noise in III-V and SiGe-based HBTs: comparison, trends, numbers”, (inv. paper), Proc. SPIE Int. Symp. on Fluctuations and Noise, Maspalomas (Spain), Vol. 5470, pp. 151-163, 2004.
- [59] J. Herricht, P. Sakalas, M. Schröter, P. Zampardi, Y. Zimmermann, F. Korndorfer, and A. Simukovic, “Verification of  $\pi$ -equivalent circuit based microwave noise model on AIII/BV HBTs with emphasis on HICUM”, Proc. International Microwave Symposium, Long Beach, California, pp. 1419-1422, 2005.
- [60] B Gu and L. Lemaitre, “Recent advancements on ADMS development”, Proc. International NanoTech Meeting, Workshop on Compact Modeling, Boston (MA), pp. 787-790, 2008.
- [61] M. Mierzwinski, P. O’Halloran, B. Troyanovsky, K. Mayaram, R.W. Dutton, “New Capabilities for Verilog-A Implementations of Compact Device Models”, Proc. International NanoTech Meeting, Workshop on Compact Modeling, Boston (MA), pp. 187-190, 2004.
- [62] Production HICUM parameters have been available at least (to the best knowledge of the authors) at the following companies or foundries: Atmel, IBM, IHP, Infineon, Jazz Semi, National Semi, ST, Toshiba, TSMC. The number of their customers (design houses) using HICUM is unknown to the authors.
- [63] M. Schröter, “High-frequency circuit design oriented compact bipolar transistor modeling with HICUM”, (inv. paper), IEICE Transactions on Electronics, Special Issue on Analog Circuit and Device Technologies, Vol. E88-C, No. 6, pp. 1098-1113, 2005.

This page is intentionally left blank

## Chapter 2

# Device Modeling Overview

Reliable circuit design requires *accurate* models for both active and passive devices. Figure 2.1 exhibits for the available approaches to semiconductor device modeling a classification in terms of model sophistication (“complexity”) versus physics retained in the model formulation (“physical basis”). The latter refers to the number of physical effects and how explicitly they are taken into account. The former refers to the required amount of work for building the model, i.e. theoretical and coding work, as well as the computational effort for its numerical evaluation, i.e. the simulation time. Obviously, solving the carrier transport at the microscopic level via the *Boltzmann transport equation* (BTE) provides the highest level of physics contents (among the approaches outlined in Fig. 2.1), but at the expense of large simulation time even just for a single device structure. For the latter, moments of the BTE in form of lower dimensional differential equations for *hydro-dynamic* (HD) or even more simple *drift-diffusion* (DD) transport are widely being used. However, both the computational and coding effort of such numerical device simulators is still prohibitive for circuit design.

A further significant simplification of the transport equations leads to explicit formulations of the charges within the device and the terminal currents in dependence of the terminal voltages. Such simplified models that are suitable for network (i.e. circuit) analysis are called *compact* models. They enable the simulation and optimization of circuits in acceptable time and can be subdivided into two classes. Semi-physical or physics-based compact models one hand employ not only analytical formulations that are based on simplified solutions of the underlying differential equations also

used in device simulation but also an equivalent circuit closely related to the physical device structure. As a consequence, the formulations and the corresponding model parameters have a physical meaning. On the other hand, table-based and polynomial compact models employ tables with data values or polynomial functions that are fitted directly to data obtained from terminal measurements of a specific device. Therefore, these models typically have very simple equivalent circuits and either no model parameters or parameters without any physical meaning. As a consequence, these models cannot be used for (i) matching analysis in, e.g., precision circuits; (ii) statistical and predictive modeling and circuit design; (iii) circuit optimization by proper device sizing. Moreover, the measurement effort for providing sufficient data over bias, temperature, and frequency can be significant, and becomes prohibitively large if models for many different device sizes are required.

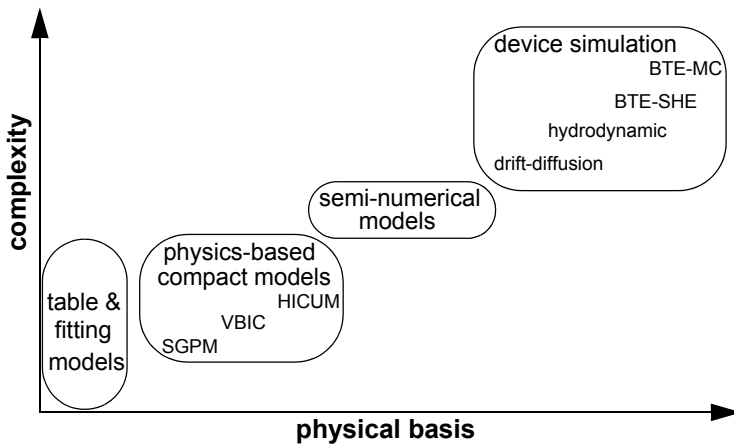


Fig. 2.1: Trade-off between complexity and physical basis of device models. See Table of Abbreviations for the acronyms.

This chapter starts with a brief overview on bipolar transistor technology in order to provide a feel for the device structures to be modelled. Then the basic set of semiconductor equations used in this book are summarized. Next, a definition of compact models is given and the relevant requirements are discussed. Finally, some definitions used for charge calculations



from numerical simulations are provided that will be used throughout the book.

## 2.1 A brief history on bipolar transistor technology

Regarding an overview on the history of bipolar technology development the reader is referred to the numerous papers that can be found in the literature (e.g. [1, 2, 3, 4, 5]). Instead, below the most important process innovations will be briefly reviewed that led to the device design of advanced transistors found today. The physical and electrical effects mentioned will be explained in more detail throughout this book.

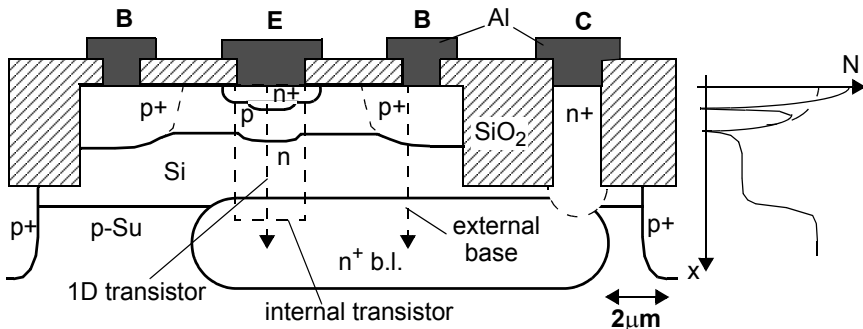


Fig. 2.2: Conventional planar bipolar transistor. Left side: schematic cross-section with 1D transistor (dashed line with arrow along the emitter center), internal transistor (dashed box under emitter window) and external base (dashed line with arrow under base contact). Right side: sketch of typical net doping profile under the emitter (solid) and external base (dashed). The abbreviation b.l. refers to buried layer.

The starting point is a conventional device structure from the mid 70ies as shown in Fig. 2.2. The mono-silicon emitter and base region were contacted directly with metal or silicide, preventing shallow junctions due to, e.g., possible “pipes”. As emitter doping, phosphorus (P) was used since its diffusion constant is similar to that of boron (B), which is needed for base doping. However, the larger diffusion constant of P versus arsenic (As) makes it difficult to achieve a shallow emitter junction, even without metal contact. Also, large P concentrations enhance B diffusion, leading to a larger base width and BC junction depth. Further issues plaguing this structure are (i) the large footprint caused by the required contact separa-

tions, (ii) a large BC junction capacitance due to the large BC-junction area and, (iii) in particular, the junction isolation. The result of the large junction widths and lateral dimensions are large parasitics that slow down the transistor speed.

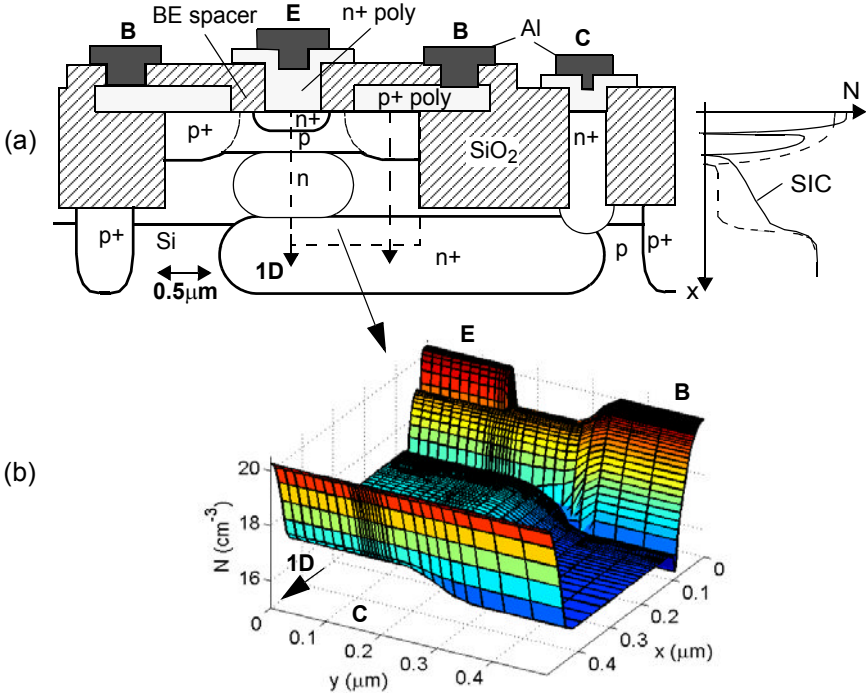


Fig. 2.3: Double-poly self-aligned bipolar transistor structure with junction isolation: (a) schematic cross-section (the dashed box defines the internal transistor) with (on the right) sketch of typical net doping profile under emitter (solid) and under external base (dashed). (b) Doping profile in the dashed box used for 2D device simulation.

In the late seventies, attempts were undertaken to overcome the above mentioned stumbling blocks and to improve speed and power dissipation. First, the poly-silicon base was introduced for contacting the Si base region and providing the dopants to create a low-ohmic external base region. The emitter opening in the poly-base also enabled a self-aligned structure, in which the emitter window and the poly-base were separated only by a relatively thin oxide called *BE spacer* or *base link* region. This way the low-

ohmic external base region could be brought much closer to the internal transistor leading to a significant reduction in base resistance. Shortly afterwards, it was found that replacing the metal emitter by poly-silicon gave superior electrical results by allowing to form shallow mono-silicon emitters without impacting speed and current gain. The resulting double-poly-silicon self-aligned structure is shown in Fig. 2.3.

In the early eighties, also the first deep-trench isolation (DTI) was developed so that it was possible to significantly reduce not only the footprint but also the cross-talk between devices. It should be mentioned though that even today DTI is often eliminated from production processes due to cost reasons. Another very important innovation was the introduction of the selectively implanted collector (SIC), usually performed self-aligned through the open poly-base or emitter window. As a consequence, the behavior of only the internal transistor can be influenced while especially the BC (feedback) capacitance of the external region can be kept to a minimum. An early idea for this implant was to reduce the internal base width by placing a small region with larger collector doping than the epitaxial doping close to the BC junction under the emitter. However this approach does very little to alleviate the collector related high-current effects (often called Kirk-effect or base push-out) which limit peak  $f_T$ . Modern transistors often use multiple SIC implants to achieve an internal collector profile shape that is optimal for a desired electrical behavior.

Despite all these measures the ideality of the DC characteristics and the speed of BJTs can only be improved up to a certain limit. For instance, making the base region thinner requires a more than proportional doping increase due to the corresponding mobility reduction. The result is only a slight increase in speed at the cost of a significant reduction in current gain and increase in tunneling current (or reduction in breakdown voltage). Similarly, an increase in collector doping leads to an increase in  $f_T$  but to a reduction in collector breakdown voltage (e.g.  $BV_{CEO}$ ). Furthermore, the lateral dimensions need to be reduced along with the vertical dimensions to achieve optimum performance. A smart approach to drastically reduce the footprint was to leverage the lithography tools of CMOS technology by integrating bipolar transistors into existing CMOS processes [5]. The added benefits of such BiCMOS technology generally are higher chip functionality and lower power dissipation, as high-frequency

bipolar front-end circuits can be combined with lower frequency baseband and large digital blocks.

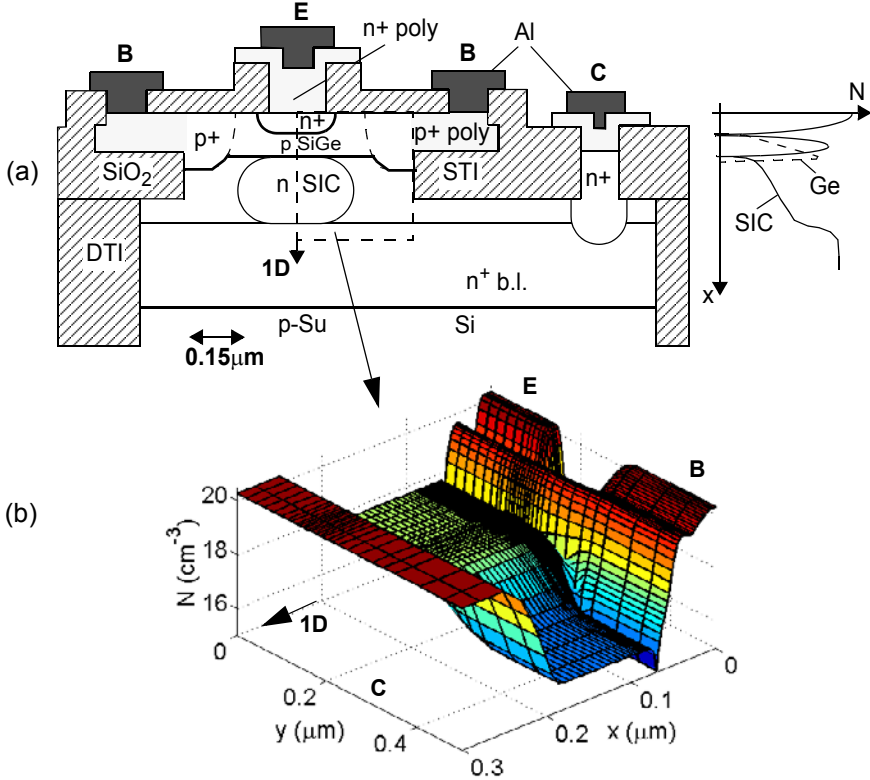


Fig. 2.4: SiGe HBT with deep trench isolation: (a) schematic cross-section with (on the right) sketch of typical doping and Ge profile. (b) Doping profile in the dashed box used for 2D device simulation.

Traditionally, bipolar transistors have maintained a speed advantage over MOS transistors of about a factor three (cf. chapter 12). In order to justify the higher cost of BiCMOS technology, and since footprint reduction alone cannot maintain that speed advantage further improvements of the bipolar transistor architecture are mandatory. This has been accomplished by both the use of atomic layer epitaxy for growing the base layer and the introduction of germanium (Ge). Initially the germanium, which has a lower bandgap than silicon, was graded across the base layer in order to create an additional drift field for the injected minority carriers, thus in-

creasing speed and Early voltage. In addition, the reduced bandgap in the base led to a higher current gain. The typical structure of the resulting SiGe heterojunction bipolar transistor (HBT) is shown in Fig. 2.4.

Since Ge and Si have a different the lattice constant the SiGe layer thickness is limited for a given Ge mole fraction [6]). This was a major reason for incorporating Ge only in the base layer, while keeping a Si collector region. As a consequence, the SiGe HBT is a so-called double heterojunction structure as compared to most III-V HBTs, which have a wide bandgap emitter with a fairly abrupt emitter base heterojunction and a narrow gap base and collector region. It is interesting to note that early SiGe HBTs did not have an emitter base heterojunction at all but, for the film stability reasons mentioned, a base-collector heterojunction. Since the latter usually causes a barrier for the main carriers (i.e. electrons in an npn transistor) it is generally considered as undesired. However, in the SiGe system this barrier only occurs at high current densities, while initially at low current densities it resides in the valence band and prevents minority injection into the collector at not too high current densities. This leads to a slightly increased  $f_T$  but also to a sharper drop towards high current densities. Advanced SiGe HBTs also contain a base-emitter heterojunction. The resulting barrier in the valence band significantly reduces hole injection into the neutral emitter and, hence, the corresponding emitter storage time, contributing to a significant increase in  $f_T$ .

The biggest delay contributions in advanced vertical SiGe HBT structures are caused by the BE injection barrier for the main (transfer) current and the finite velocity through the BC SCR. The latter can be reduced by shorter SCR width and utilizing non-local carrier transport with its high velocity, while the former requires a reduction of the BE depletion capacitance and an increase of the transconductance.

Presently existing SiGe HBTs in 0.18 $\mu\text{m}$  or even 0.25 $\mu\text{m}$  BiCMOS mainstream production technologies achieve their peak ( $f_T, f_{max}$ ) around (150,180)GHz (@  $V_{BC}=0\text{V}$ ) with  $BV_{CEO}$  values around 2.2V (see [1] for an overview). Processes with peak speed of (210,250) GHz at  $BV_{CEO}$  around 1.8V are close to production (e.g. [7, 8, 9, 10]). So far, the highest combination of ( $f_T, f_{max}$ ) values measured are around (300,350)GHz at  $BV_{CEO}$  around 1.5V (e.g. [11, 12, 13]). A record speed of  $f_T=500$  GHz was obtained at 77 K [14] and of 620 GHz at 35 K [15], both at moderate

$f_{max}$ . Although most process technologies are advertised by quoting their high performance transistor type, they also offer usually at least one other transistor version in order to enable successful circuit designs. In all cases, a “high-voltage” version is available in which speed ( $f_T, f_{max}$ ) is traded in for a higher collector breakdown voltage ( $BV_{CEO}$ ). Often, also a “medium performance” version is offered as a compromise between speed and breakdown or tuned for a particular target product application. These options are realized with just one different mask per version for the selectively implanted collector.

The idea of HBTs [16, 17] was realized already in the seventies [18, 19]. However, this path of HBT development has been based on bulk material compositions from the III-V group (e.g. AlGaAs, InGaAs), hoping to be able to utilize the higher electron mobility for significant speed improvement. The lack of an efficient passivation and many other material related issues plagued the development and, in particular, the commercial deployment of these technologies for a long time. Since the nineties though, circuits with reasonable complexity have been built and integrated into products. GaAs HBTs presently dominate the cell phone power amplifier market. Record speed beyond 500 GHz has been demonstrated with InP HBTs by different groups (e.g. [20–22]). A main disadvantage of all III-V technologies is their fairly low integration limit (as compared to Si-based BiCMOS), preventing cost efficient system solutions for medium and large volume applications. However, III-V technologies are often excellent choices for small volume niche markets.

Throughout this book model development and the related assumptions as well as model verification are supported by device simulation. The vertical doping profiles under the emitter employed for 1D analysis (cf. Fig. 2.3 and 2.4) are presented below, along with a brief discussion of the selection process that is based on the technology development overview given above. The terminal nodes of the 1D transistor are designated by primed indices, and so are the corresponding voltages, i.e.  $V_{B'E'}$ ,  $V_{B'C'}$ .

For understanding the fundamental transistor operation and the main physical effects, the BJT with a doping profile shown in Fig. 2.5a has been selected. Such a profile is suitable for deriving the basic theory in Chapter 3. In order to maintain a reasonable current gain the emitter concentration usually reaches the solubility limit, while the base is moderately

doped. A constant (epitaxial) collector doping concentration is one of many choices for the collector profile; it was selected here since it is suitable for explaining the important effects and for illustrating the assumptions made during model development. At the end of the epitaxial collector the concentration increases toward the highly doped buried layer. In addition to the doping the equilibrium carrier concentrations are inserted.

The electrical DC behavior of the 1D BJT is characterized by the collector and base current as function of the BE voltage. (This diagram is often called Gummel plot.) According to Fig. 2.5b an (almost) ideal behavior of the collector current density  $J_C$  is observed up to a certain critical current density  $J_{CK}$ , above which the increase of current and associated transconductance with  $V_{B'E'}$  are significantly reduced. This is caused by so-called high-current effects in the collector. In contrast to  $J_C$  the base current density  $J_B$  continues to increase unchanged beyond  $J_{CK}$ .

An important figure of merit for characterizing the small-signal high-frequency behavior is the transit frequency  $f_T$ . Like in Fig. 2.5c, it is typically shown as a function of collector current *density* in order to allow (i) a comparison between different transistor types and sizes and (ii) the selection of the proper size for circuit design. The decrease of  $f_T$  at low  $J_C$  is mainly caused by depletion capacitances and parasitic capacitances between the terminals. (Note that the latter do not exist in a 1D transistor.) The fall-off towards high  $J_C$  is caused by the already mentioned high-current effects, which lead to a strong increase in minority charge with current density. The corresponding storage time of the overall transistor,  $\tau_S$ , has been inserted in Fig. 2.5c to provide a feel for the effect. Around the critical current density, which is indicated by the arrow,  $\tau_S$  increases very rapidly. For circuit design, it is important to describe the begin of the increase accurately. Once  $\tau_S$  has increased by more than a factor 3 to 5, just the rough behavior needs to be captured by a model.

The selected BJT is representative for modern BJT technologies, which still find many cost-efficient circuit applications in linear and power products. In these areas, often complementary vertical pnp transistors have been developed with even quite balanced performance w.r.t. the npn type.

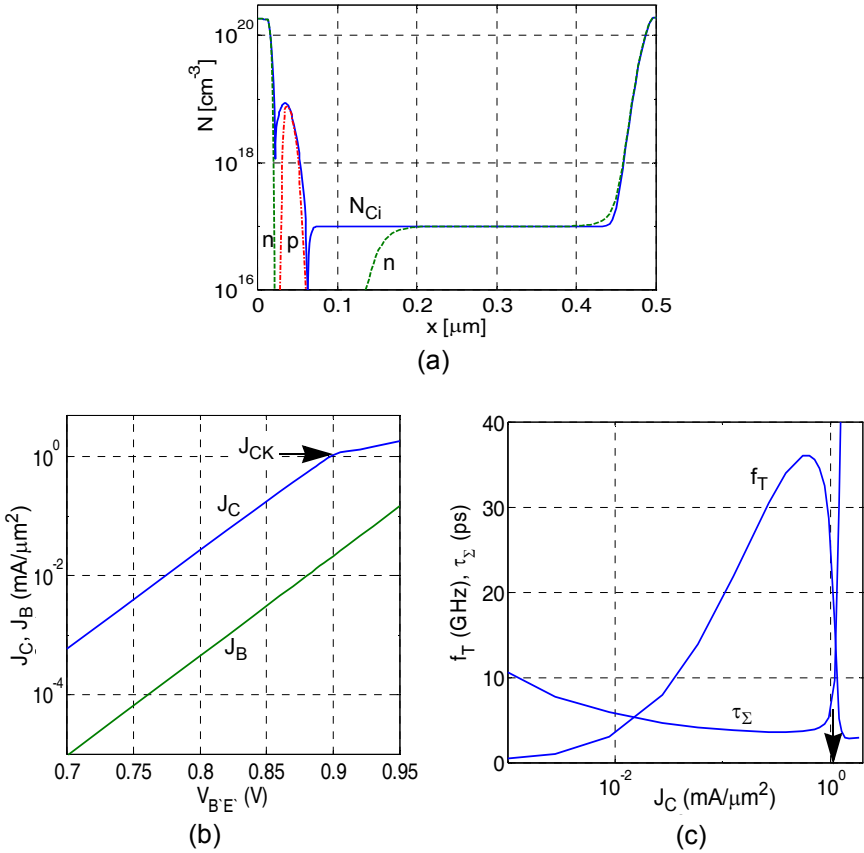


Fig. 2.5: Selected Si-BJT: (a) vertical (1D) net doping profile (solid line) with equilibrium carrier densities (dashed lines); (b) Gummel characteristics; (c) transit frequency  $f_T$  and Regional Approach transit time  $\tau_\Sigma$  vs. collector current density.  $V_{B'C'} = 0$  V,  $T = 300$  K. The length of the simulated region in (a) is defined as  $L_X$ .

Modern high-speed BiCMOS technologies are realized with SiGe HBTs. Figure 2.6a displays the doping and composition profile of the npn HBT employed in this book for deriving and discussing the corresponding theory. Again, a spatially constant collector doping is used, but now the concentration is higher and the internal collector width is smaller. Both these measures contribute significantly to a speed increase of the 1D structure compared to the BJT. The other main portion of the speed increase results from the introduction of a graded Ge concentration and the associated



aiding drift field for electrons across the (neutral) base region. A small Ge step of about 2% at the BE junction causes a barrier for hole back injection into the emitter. The resulting higher current gain has then been traded for the higher base doping, which in turn reduces the base resistance and thus the delay time of a (2D and 3D) structure. Note that compared to the BJT the base peak concentration is shifted away from the BE junction. This in combination with a smaller slope of the emitter profile leads to a reduction of both the depletion capacitance and leakage current as compared to a BJT structure. Since the BE doping profile of the HBT in Fig. 2.6a is still similar to that of BJTs, the structure is often referred to as conventional emitter doping (CED) HBT. This is in contrast to the originally conceived low emitter concentration (LEC) HBT.

The corresponding Gummel characteristics are shown in Fig. 2.6b. Compared to the BJT a  $V_{B'E'}$  shift of about -30mV can be observed. Furthermore, towards high current densities the increase of  $I_C$  drops much more abruptly, and the heterojunction barrier resulting from the sharp drop of the Ge in the base-collector region can cause  $I_B$  to increase even faster than ideal. The kink in  $I_B$  at about 0.88V is due to impact ionization which was calculated here using a non-local breakdown model.

The HF performance of the selected HBT structure is shown in Fig. 2.6c along with the corresponding total storage time. A much sharp drop of  $f_T$  and increase of  $\tau_S$  towards high current densities can be observed due to the same reasons mentioned earlier. It is interesting to note that the onset of high-current effects can still be quite well calculated by the critical current density  $J_{CK}$ , which is also indicated in Fig. 2.6b and c.

The selected structure is a compromise w.r.t. various aspects. Although the peak of  $f_T$  is not as high as for the most advanced SiGe HBTs the structure is a reasonable representation of mainstream SiGe HBT technology. On the other hand, it is suitable for deriving the necessary theory, and the results of DD simulation are still quite accurate.

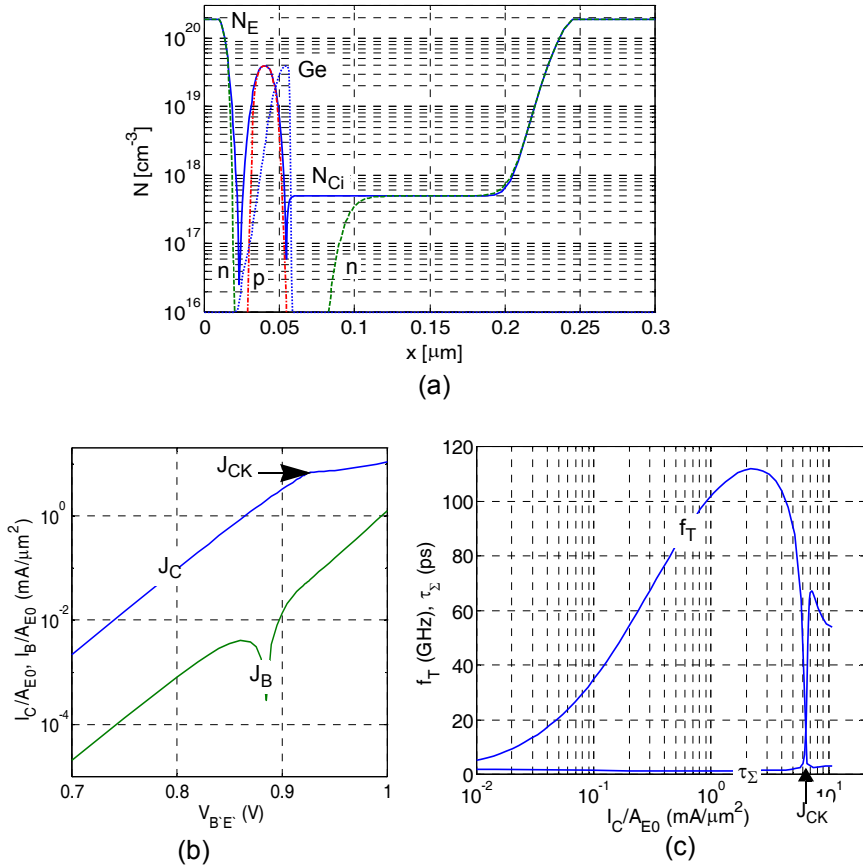


Fig. 2.6: Selected SiGe HBT: (a) vertical (1D) net doping profile (solid line) with equilibrium carrier densities (dashed lines) and Ge profile (dotted line); (b) Gummel characteristics; (c) transit frequency and Regional Approach transit time vs. collector current density.  $V_{B'C'} = 0V$ ,  $T = 300K$ . The length of the simulated region in (a) is defined as  $L_x$ .

Table 2.1 summarizes the most important technological data of the transistors used in this book.  $x_{jE}$  ( $x_{jC}$ ) is the emitter (base) junction depth;  $w_{Bm}$  is the metallurgical base width;  $N_{Ci}$  is the (epi) collector doping concentration under the emitter and  $w_{Ci}$  is the corresponding (epi) collector width. The latter is defined at the location where the collector doping has increased to  $2 N_{Ci}$  since from this point on the resistance increase towards

the buried layer peak is negligible. The values for the zero-bias internal base sheet resistance are  $6.4 \text{ k}\Omega/\text{sq}$  for the BJT and  $1.6 \text{ k}\Omega/\text{sq}$  for the HBT.

The power transistor versions have the same emitter, base and buried layer doping profile as the high-speed versions, but a lower collector doping and, in case of the HBT, also a wider collector in order to sustain higher breakdown voltages. Their results will be used whenever differences between high-speed and power transistors need to be demonstrated and clarified.

transistor	$x_{jE}/\mu\text{m}$	$x_{jC}/\mu\text{m}$	$w_{Bm}/\mu\text{m}$	$w_{Ci}/\mu\text{m}$	$N_{Ci}/\text{cm}^{-3}$
BJT	0.020	0.063	0.043	0.38	$1 \cdot 10^{17}$
power BJT	0.022	0.067	0.045	0.375	$2 \cdot 10^{16}$
HBT	0.024	0.054	0.03	0.15	$5 \cdot 10^{17}$
power HBT	0.024	0.055	0.031	0.25	$2 \cdot 10^{16}$

Table 2.1: Most important technological parameters of the transistors used.

## 2.2 Semiconductor equations

For understanding device operation and improving their performance, carrier transport in semiconductors is one of the most important topics (e.g. [23]). Each mobile carrier in a semiconductor is characterized by its location in *real* space and its momentum in *phase* space. Both location and momentum can change over time due to forces that act upon the carriers. The most simple form of viewing the carrier ensemble is a charged gas of free carriers, that is influenced by external forces, such as electric and magnetic fields or mechanical stress. At the microscopic level, one could try to follow the path of each carrier in a semiconductor volume under the influence of external forces and scattering. In order to do this, fundamental laws of conservation and carrier motion need to be formulated. The most often used fundamental equation for describing microscopic carrier transport in semiconductors is the BTE, which for realistic doping profiles can only be

solved numerically. The Monte-Carlo (MC) method is the most widely employed solution technique (e.g. ([24])), in which the movement of each carrier is simulated as it undergoes the various interactions discussed above. The resulting solution yields the statistical distribution of carriers in real and phase space.

Unfortunately, the MC method is computationally very expensive and also suffers from an insufficient resolution of low carrier densities, such as minority carriers in a bipolar transistor at low injection. Therefore, more efficient solution methods have been sought. A significant improvement regarding both computational effort and carrier density resolution has been obtained by employing a deterministic approach to solving the BTE via a spherical harmonics expansion (SHE) [25]. However, the simulation time is still relatively large, especially for the 2D case. In most of today's technologies, carrier transport involves a *large* number of particles. Hence, for most device applications, it is sufficient to consider macroscopic (localized) quantities such as the *average* carrier density, current density or energy density, which can be described by a simplified set of equations. Such equations are called moments of the BTE (e.g. [26]). Taking just the first two moments leads to the well-known *drift-diffusion* formulation (DDF), while taking one or two additional moments for the energy and its flux yields various sets of *hydro-dynamic* formulations (HDF). Hence, a hierarchy of equations exists that can be adapted to the respective simulation problem ([27]). The solution of these equations is called (*numerical*) *device simulation*.

Despite its simplified form, the DDF is still computationally too expensive for circuit design and, thus, cannot be used in compact models. However, the DDF can be employed for developing compact models, and its (numerical) solutions can serve as reference for compact model verification. Therefore, the most important equations constituting the DDF are summarized below (cf. [28, 29, 23, 2]).

For the vast majority of semiconductor devices the influence of magnetic fields can be neglected. This permits to express the electrostatic field by a gradient of a scalar electrostatic potential  $\psi$ ,

$$\vec{E} = -\text{grad}\psi. \quad (2.1)$$

and, in turn, allows *Poisson's* equation to be written as

$$\text{div}(\varepsilon \text{ grad} \psi) = -\rho \quad (2.2)$$

with  $\varepsilon = \varepsilon_0 \varepsilon_s$  as spatially dependent permittivity, where  $\varepsilon_s$  is the relative permittivity of the respective semiconductor. Furthermore,

$$\rho = q(N_D^+ - N_A^- + p - n) \quad (2.3)$$

is the space-charge density, consisting of the ionized donors ( $N_D^+$ ) and acceptors ( $N_A^-$ ), as well as of the mobile carriers, namely the electron density

$$n = n_{ir} \exp\left(\frac{V_n}{V_T}\right) \exp\left(\frac{\psi - \phi_n}{V_T}\right), \quad (2.4a)$$

and hole density

$$p = n_{ir} \exp\left(\frac{V_p}{V_T}\right) \exp\left(\frac{\phi_p - \psi}{V_T}\right), \quad (2.4b)$$

with  $n_{ir}$  as the intrinsic carrier density of a reference material;  $\phi_n$  and  $\phi_p$ , respectively, is the quasi-Fermi potential of electrons and holes, respectively. Furthermore, these formulations already include all effects that cause the band edges to depend on temperature, doping and material composition via the so-called band potentials  $V_{n,p}$  for each carrier (e.g. [30]):

$$V_p = -\frac{\chi_0 - \chi_r}{q} + V_T \ln\left(\frac{N_V}{N_{Vr}}\right) - \frac{W_{g0} - W_{g0r}}{q} + \frac{\Delta W_V}{q}, \quad (2.5a)$$

$$V_n = \frac{\chi_0 - \chi_r}{q} + V_T \ln\left(\frac{N_C}{N_{Cr}}\right) + \frac{\Delta W_C}{q}. \quad (2.5b)$$

Taking the sum of the band potentials gives

$$V_n + V_p = V_T \ln\left(\frac{N_C}{N_{Cr}} \frac{N_V}{N_{Vr}}\right) - \frac{W_{g0} - W_{g0r}}{q} + \frac{\Delta W_{hd}}{q} + \frac{\Delta W_m}{q}. \quad (2.6)$$

with  $\Delta W_{hd}$  and  $\Delta W_m$ , respectively, as the *total* (measurable) bandgap-variation due to high-doping effects and material composition, respectively. For homogeneous semiconductors only the contribution  $\Delta W_{hd}$  remains. In inhomogeneous material such as heterojunction, the effective mass also changes with composition position and so do the effective densities of states. The introduction of the band potentials allows to keep using carrier density formulations in device simulation that are very similar to those

from Boltzmann statistics instead of having to use more complicated Fermi-statistics.

For semiconductor device simulation and modeling, often a generalized *effective* intrinsic carrier concentration is used, which follows from the *pn* product at equilibrium

$$n_i^2 = n_{ir}^2 \exp\left(\frac{V_n + V_p}{V_T}\right). \quad (2.7)$$

Carrier dynamics are modeled by balance and transport equations. The first moment of the BTE leads to a balance equation that is known as the *carrier continuity* equation for holes and electrons, respectively:

$$\text{div} \vec{J}_p = -q \left( R + \frac{\partial p}{\partial t} \right), \quad (2.8a)$$

$$\text{div} \vec{J}_n = q \left( R + \frac{\partial n}{\partial t} \right), \quad (2.8b)$$

Note, that these equations can also be derived from Maxwell's equations.

Under the assumption of classical carrier drift and diffusion as the only transport mechanisms, the equations for the current densities read for the general three-dimensional case

$$\vec{J}_p = -q\mu_p V_T \text{grad } p + qp\mu_p \vec{E}_p, \quad (2.9a)$$

$$\vec{J}_n = q\mu_n V_T \text{grad } n + qn\mu_n \vec{E}_n, \quad (2.9b)$$

$\mu_c$  ( $c = p, n$ ) is the (macroscopic) carrier mobility that is described as a function of the *local* field, given by  $\text{grad}(\varphi_c)$ , as

$$\mu_c = \frac{\mu_{c0}}{\left[ 1 + \left( \frac{\text{grad}(\varphi_c)}{E_{lim,c}} \right)^{\beta_{\mu c}} \right]^{1/\beta_{\mu c}}} \quad (2.10)$$

with the low-field value  $\mu_{c0}$  and  $\beta_{mc}$  as parameter. The critical field  $E_{lim,c} = v_{s,c}/\mu_{c0}$  is given by the saturation velocity  $v_{s,c}$  that is reached for high fields. Therefore, the drift velocity as a function of field reads

$$v_c = v_{s,c} \frac{\text{grad}(\varphi_c)/E_{lim,c}}{\left[1 + \left(\frac{\text{grad}(\varphi_c)}{E_{lim,c}}\right)^{\beta_{\mu c}}\right]^{1/\beta_{\mu c}}}. \quad (2.11)$$

The current densities are also a function of an effective electrical field,

$$\vec{E}_p = -\text{grad}\psi + \text{grad}V_p = \vec{E} + \text{grad}V_p, \quad (2.12a)$$

$$\vec{E}_n = -\text{grad}\psi - \text{grad}V_n = \vec{E} - \text{grad}V_n, \quad (2.12b)$$

which consists of the gradient of not only the electrostatic potential  $\psi$  (see (2.1)) but also of the band potentials.

Above equations cover the most important basic transport mechanisms in existing Si BJTs and SiGe HBTs. However, there are various effects that can be found in more advanced devices but cannot be described by the DDF:

- Non-equilibrium (quasi-)ballistic transport, which becomes important in Si/SiGe bipolar transistors and in most III/V HBTs, respectively, at base widths below approximately 40nm and 100nm, respectively.
- Thermionic emission over barriers.
- Effects caused by mechanical stress of the lattice. However, the impact of strain from material composition usually is empirically taken into account within, e.g., the carrier mobility formulation.
- Quantum effects such as lateral or vertical confinement and tunnelling through barriers.

Non-equilibrium transport can be described to first order by solving an energy balance equation. More or less sophisticated formulations of this equation and also the required energy transport equation exist in the literature (e.g. [27]), all being usually referred to as hydrodynamic equations. For compact modeling, only a simplified version is suitable, which can be written as

$$\frac{dT_c}{dx} + \frac{T_c - T_L}{\lambda_c} + \frac{2q}{5k_B} E_c = 0. \quad (2.13)$$

$T_c$  is the carrier temperature representing its kinetic energy,  $T_L$  is the lattice temperature,  $E_c$  is the carrier related electric field (i.e. the gradient of

respective quasi-Fermi potential), and  $\lambda_c$  is the carrier energy relaxation length. The general solution of the above equations reads

$$T_c(x) = T_L + \frac{2q}{5k_B} \int_0^x E_c(\xi) \exp\left(\frac{\xi - x}{\lambda_c}\right) d\xi, \quad (2.14)$$

which corresponds to a spatial weighting of the field and its impact on carrier energy.

In summary, the DDF as the most simple version of the semiconductor basic equation system consists of the three fundamental equations (Poisson's equation and the carrier continuity equations), which are coupled with each other by equations for material properties, such as permittivity, carrier densities and carrier transport. Overall, these equations represent a system of coupled non-linear, partial differential equations in 3D space and in time.

In this book, device simulation is used extensively for compact model development. In most cases, a 1D solution is sufficient due to the 1D nature of the intrinsic BJT/HBT structure, and DDF is sufficient to visualize the fundamental behavior. However, where applicable also 2D solutions are shown. In all cases, the simulator DEVICE [31] is used the physical parameters of which have been calibrated (e.g. [32]) based on measured transistors and comparisons with other simulators (e.g. [33, 34]). Note, that HDF simulation requires a proper weighting of not only the  $\text{grad}(T)$  term in the carrier transport formulation but also of the energy flux formulation as well as a careful device type-specific (i.e. MOS or bipolar transistor) tuning of the energy relaxation time [35]. Figure 2.7 shows a comparison between DDF, HDF, and SHE solution of the BTE for the SiGe HBT used in this book. Both BTE and HDF show a somewhat less ideal collector current characteristic and visibly higher transit frequency than DDF. These results are similar to those in [37] obtained for a LEC doping profile. The differences increase with decreasing vertical transistor dimensions.



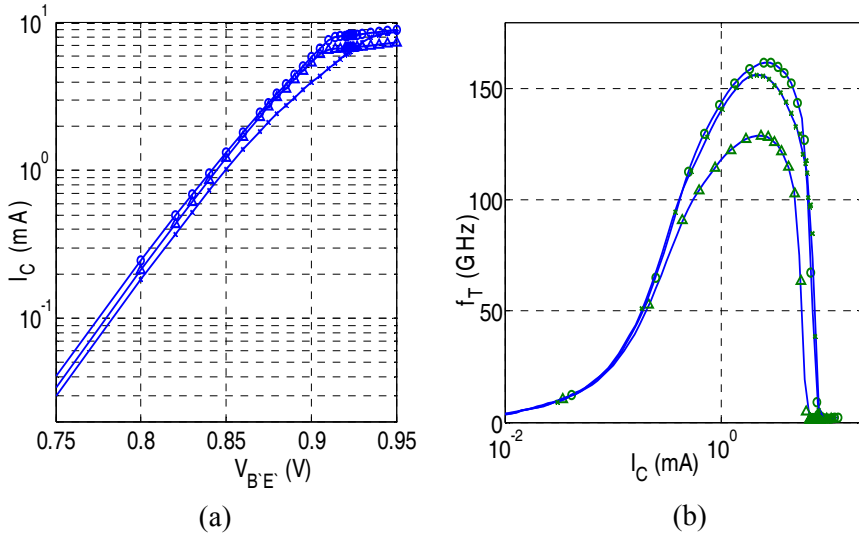


Fig. 2.7: Comparison between DDF ( $\Delta$ ), HDF (o), and SHE solution of the BTE (x) for the CED HBT: (a) DC collector current and (b) transit frequency [36].

### 2.3 Compact modeling

The main goal of compact models is to enable circuit design. Since circuits generally consist of several up to millions of transistors, compact models need to be sufficiently simple from a computational point of view in order to ensure acceptable simulation times. In addition to the demand for simplicity, there exist quite a number of additional requirements for a production suitable compact model, which are summarized in Fig. 2.8 and discussed below.

Modern technologies with their increasingly complex design tasks and complicated device physics have led to specialized expertise in these areas. This has made it difficult to develop compact models that can be easily understood by non-specialists. Hence, the observed trend is that the general circuit design community demands models, which accurately cover a wide bias and temperature range, relieving the designer as much as possible from knowing the validity limits of a model. Another important require-

ment is the capability of (preferably smooth) geometry scaling for automated circuit optimization purposes.

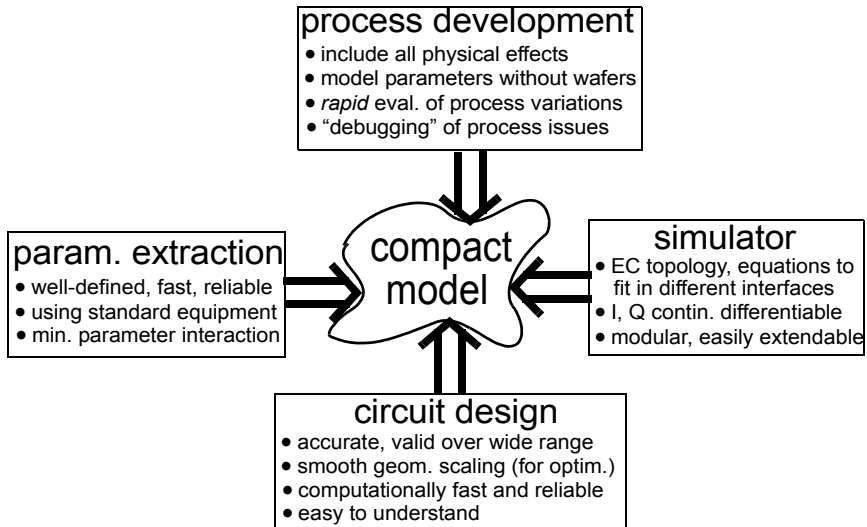


Fig. 2.8: Main requirements for an industrially and production suitable compact model.

Model reliability in circuit simulators is closely related to the model formulation. Good convergence behavior requires continuous derivatives of currents and charges. Furthermore, the more modular a model is formulated the faster and less error-prone extensions can be integrated. Also, the simulation time in many simulators increases with the number of nodes, i.e. the complexity of the equivalent circuit.

The best model cannot be used without properly determined parameters. A key task for making a compact model successful thus is to establish a well-defined parameter extraction methodology and infrastructure which is greatly aided by properly formulated model equations and well-defined parameters (and test structures). Ideally, from a parameter extraction point of view, model parameters should not be correlated; however, the demand for simplicity and explicit solutions may not always allow this. Standard measurement equipment and procedures should be sufficient for parameter extraction related data acquisition.

Process development today is heavily oriented towards certain target applications. As a consequence, it is highly desirable to be able to predict cir-

cuit performance for a new process as early as possible. In fact, the trend is to co-develop process and circuits with the goal to have working samples already at the time of process qualification. This so-called concurrent engineering obviously contributes to a significant reduction in time-to-market. The corresponding cost reduction is achieved if compact models are capable of accurately predicting circuit performance early in the process development phase. Such models obviously must have a strong physical basis and, thus, must also allow to include the statistical variations of a mature process. Finally, together with the proper test structures, it is even desirable (and in fact possible) to use the model formulations for process debugging.

The above requirements are best met by a *physics-based compact model*, which can be characterized by the following properties:

- The device structure is generally described by an equivalent circuit (EC) containing lumped elements for the various or most important spatial regions.
- Each element value as a function of, e.g., bias and temperature is calculated by computationally as inexpensive as possible equations formulated as closed-form analytical solutions, preferably in explicit form. In other words, numerical solutions of differential equations are usually prohibited, thus excluding numerical device models.
- The element equations contain a reasonably small set of model parameters, which can be determined from electrical measurements at the device terminals. In other words, it is undesirable (e.g. from a foundry intellectual property point of view) to have to know or to provide *internal* structural information of a device (such as doping profiles) for determining model parameters. However, it is desirable for predictive and statistical modeling if parameters can be expressed in terms of such information.

Above properties can be summarized formally by writing the value of the  $k^{\text{th}}$  EC element,

$$u_k = f(\{\mathbf{i}, \mathbf{v}\}, T_j, \mathbf{d}, \mathbf{e}, \mathbf{t}, \sigma), \quad (2.15)$$

with the following variables:

- $\{\mathbf{i}, \mathbf{v}\}$  denotes the bias point;
- $T_j$  is the junction temperature;

- $\mathbf{d}$  is the vector containing information on the dimensions of the device structure (design rules, layout) and device *configuration*  $\mathbf{c}$  (defined by emitter width, length, number and location of contacts);
- $\mathbf{e}$  represents specific electrical parameters (such as sheet resistance, capacitance per unit area or length) and also includes physical parameters (such as mobility), usually though in merged form;
- $\mathbf{t}$  represents technology parameters such as average doping concentrations and important vertical dimensions;
- $\sigma$  contains the process tolerance of typically a fairly small subset of variables in  $\mathbf{d}$  and  $\mathbf{e}$  or, alternatively,  $\mathbf{t}$  (cf. chapter on Applications).

The development of a compact model obviously requires a lot of simplifications with respect to the equations of the DDF or HDF. In fact, even today's advanced compact models are still mostly based on the DDF. In order to arrive at analytical solutions of the (coupled) semiconductor differential equations usually the vertical and lateral spatial direction are treated separately, reducing the problem for the intrinsic transistor to a 1D case (cf. Fig. 2.2). However, even with this simplification no analytical solution exists for realistic device structures (i.e. doping profiles). Hence, further partitioning both spatially and over bias is usually performed until a single semiconductor equation remains that can be solved analytically (at least in approximate form). Generally, average values for doping, field, mobility etc. need to be assumed entering, along with dimensions, the resulting model parameter expressions. Although such relations establish the physics contents of a model, a direct calculation of the respective model parameter would yield too inaccurate electrical characteristics. Therefore, these errors are compensated for by a parameter determination directly from electrical terminal characteristics. In addition to physics-based parameters, piecing together solutions from different bias regions in continuously differentiable form often requires so-called smoothing parameters that do not contain physics-based information. In other words, only a limited subset of the model parameter vector  $\mathbf{m}$  is usually related to physical parameters. The above rough description of a compact model development procedure hopefully sets the right level of expectations with respect to model use.

## 2.4 Charge definitions in bipolar transistors

Compact models require numerically efficient but accurate analytical equations for describing charge storage and static characteristics. Since such analytical formulations usually can only be obtained through simplifying assumptions of the semiconductor equations, they need to be compared to reference results from either measurements or device simulation. For instance, the designation “depletion capacitance” is commonly used in compact models but its definition is mostly based on classical theory of abrupt pn junctions which fails at a higher forward bias operation. The definition of current dependent (minority) charges is even fuzzier, and analytical descriptions are often based on ideal assumptions and then pieced together. A similar approach is often used for describing static characteristics. The problem starts when such approaches do not yield satisfactory results and have to be somehow modified or extended. At that point, clear definitions on how to calculate, e.g., charges and currents directly from device simulation are required. Therefore, the goal of this section is to provide clear definitions of important quantities used for compact modeling, guided by physical understanding of BJT and HBT operation.

For high-speed applications the accurate description of the dynamic transistor behavior, which is determined by the charge stored in the device, is most important. Since the total transistor is electrically neutral it is sufficient to consider only one type of carriers. In case of an npn transistor the change of the hole charge,  $dQ_p$ , is the appropriate choice since  $dQ_p$  represents the carriers interacting with the control electrode “base”. For *quasi-static* operation the internal state of a transistor, given by the distribution of the carrier densities, is uniquely defined by the values of the controlling terminal voltages  $V_{B'E'}$  and  $V_{B'C'}$ . With these as independent variables the small-signal charge can always be expressed as *total* differential:

$$dQ_p = \left. \frac{\partial Q_p}{\partial V_{B'E'}} \right|_{V_{B'C'}} dV_{B'E'} + \left. \frac{\partial Q_p}{\partial V_{B'C'}} \right|_{V_{B'E'}} dV_{B'C'}. \quad (2.16)$$

The goal is to find an accurate description of  $dQ_p$  as a function of bias  $\{\mathbf{i}, \mathbf{v}\}$ .

An often found approach is to partition the charge variation into a contribution from uncompensated carriers, the so-called depletion charge  $Q_j$ , and from carriers compensating each other at any given location, the so-called minority charge  $Q_m$ :

$$dQ_p = dQ_j + dQ_m. \quad (2.17)$$

This partition as well as the exact definition of its components is arbitrary though. In classical theory, the device is subdivided into space-charge regions, which are fully depleted of mobile carriers, and into neutral regions, where minority carriers are fully neutralized by majority carriers. In this so-called Regional Approach (RA), it is then assumed that  $dQ_j$  and  $dQ_m$  are restricted only to their corresponding regions. The results are the well-known formulas for, e.g., the depletion capacitances and the base transit time (cf. chapter 3).

The basic idea of the Regional Approach is very useful for developing compact models and can also be applied to device simulation, which serves as reference. However, since the general solutions obtained by device simulation do not contain abrupt boundaries between the different regions, the boundaries and regions have to be defined in a more general way. Furthermore, the distinction between depletion and minority carrier related charges becomes very difficult above certain injection levels. Hence, a definition of the relevant variables is given below and is based on a small-signal variation of the (1D) terminal voltages.

The definition starts with the change of minority (mobile) carriers

$$\delta m = \begin{cases} \delta p & , |\delta p| < |\delta n| \\ \delta n & , |\delta n| \leq |\delta p| \end{cases}, \quad (2.18)$$

with the symbol " $\delta$ " indicating the response of the carriers to a terminal voltage change (i.e.  $\delta = d/dV$ ). The corresponding change of the minority charge in the transistor is then given by

$$\delta Q_m = qA_E \int_0^{L_x} \delta m dx. \quad (2.19)$$

With the similarly given change of the hole charge,

$$\delta Q_p = qA_E \int_0^{L_x} \delta p dx, \quad (2.20)$$

the small-signal depletion charge is then automatically defined as

$$\delta Q_j = \delta Q_p - \delta Q_m = qA_E \left[ \int_0^{L_x} \delta p dx - \left( \int_{x(|\delta p| < |\delta n|)} \delta p dx + \int_{x(|\delta n| \leq |\delta p|)} \delta n dx \right) \right]. \quad (2.21)$$

The designation  $x(|\delta p| < |\delta n|)$  indicates that the integration is to be performed only within those regions in which holes are minorities, i.e. where  $|\delta p| < |\delta n|$  holds;  $x(|\delta n| \leq |\delta p|)$  has a similar meaning. Obviously, in regions where holes are minorities the first and second integral on the r.h.s. of (2.21) just cancel, leading to the final definition of the depletion charge

$$\delta Q_j = qA_E \int_{x(|\delta n| \leq |\delta p|)} (\delta p - \delta n) dx = qA_E \int_{x(|\delta p| < |\delta n|)} (\delta n - \delta p) dx. \quad (2.22)$$

The most right expression results from charge neutrality in the total transistor ( $\delta Q_p = \delta Q_n$ ) and is more suitable for calculating the depletion charges of npn transistor structures from device simulation.

#### 2.4.1 Regional analysis of bipolar transistors

For developing compact model equations, the minority and depletion charge need to be subdivided into their regional components. For the minority charges, classical transistor theory provides solutions for neutral regions, while for the depletion charges, classical theory indicates a partitioning into a BE and BC space-charge (junction) region related component  $\delta Q_{jE}$  and  $\delta Q_{jC}$ , respectively. Therefore, boundaries between a neutral region (NR) and a space-charge region (SCR) need to be defined. In contrast to the ideal conditions considered in classical theory, realistic transistors do not contain abrupt changes of carrier densities and, hence, abrupt boundaries between NRs and SCRs. The equivalent of an abrupt change of the space-charge density  $d\rho = q(dp - dn)$  at the SCR boundary in form of a delta-function is in a realistic transistor a Gaussian-shaped distribution of the differential space-charge density around the edge of the SCR as shown in Fig. 2.9(a) and (b). However, it is important to consider the space-charge variations with BE and BC short, respectively, and not

with BE and CE short, respectively. The reason for this is that at higher current densities the BC space charge density depends on the mobile charge in the BC region, which in turn depends strongly on the variation of  $V_{B'E'}$ . Therefore, a CE short - although required for the calculation of, e.g., the transit frequency - does not render useful results for determining the BC SCR *boundaries*. This becomes clearly visible from Fig. 2.9(a) which contains the dynamic space-charge density distribution at two different collector current densities, one far below peak  $f_T$  and one at peak  $f_T$ . The peaks left and right from the BC junction are uniquely defined only in the low-current curve. At high current densities (dashed lines), the  $V_{B'E'}$ -controlled modulation of the (mobile) electron charge in the BC SCR is much larger than the actual space-charge variation at the BC SCR boundaries from a  $V_{B'C'}$  change only (cf. (b)) and, hence, do not permit anymore to detect the boundaries of the BC SCR. In contrast, for  $\delta\rho(dV_{B'E'} = \text{const})$  in Fig. 2.9b the peaks at the BC junction are clearly defined also at high current densities. Note also the very different magnitude that a  $V_{B'E'}$  and  $V_{B'C'}$  change have on the carrier dynamics.

As a result of the above considerations and in order to clearly separate BE and BC voltage change related charge variations, the hole charge variation is written as (2.16) which corresponds to a first-order Taylor series of the hole charge and an instantaneous (memory-less) reaction. Thus, (2.16) holds for quasi-static (QS) operation and is consistent with the definition of  $f_T$ . The partial derivatives in (2.16) define the QS capacitances associated with the respective junction voltage.

The partitioning of the charge and the definition of the particular region boundaries is in principle arbitrary. For high-speed applications it is useful to define the boundaries based on changes of the carriers rather than based on DC carrier distributions as in [39]. An often referenced approach was given in [38], which however provides neither a separation between depletion and minority charge nor a definition of physical region widths (but only of electrical widths). A comparison of different regional analysis approaches for BJTs and HBTs is given in [40].



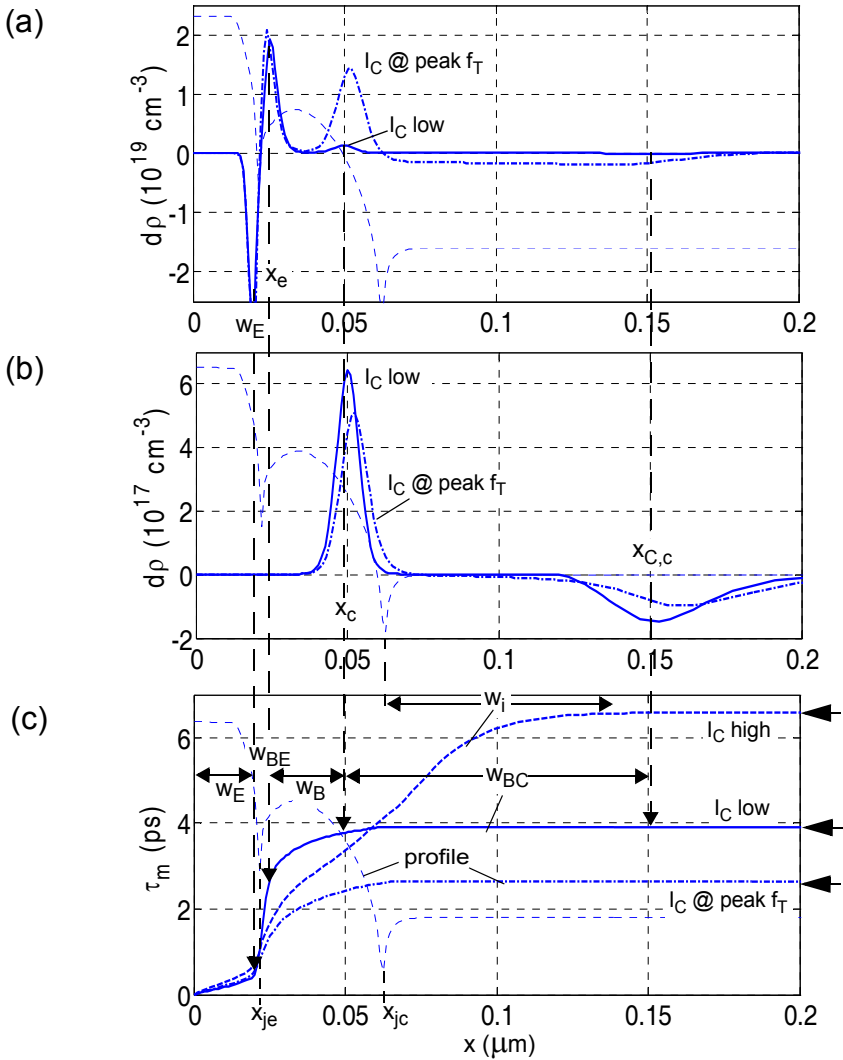


Fig. 2.9: Dynamic space-charge density distribution for a BJT at different collector current densities: (a) BE related change with CE short, (b) BC related change with BE short. (c) Accumulated transit time with regional widths inserted for the low current bias point and with just  $w_i$  for the high-current region curve. Bias currents: 0.027 mA (solid), 0.64 mA (dash-dot), 1.1 mA (dashed) at  $V_{B'C'} = 0\text{V}$ . The doping profile (dotted line) is not drawn to scale.

The SCR boundaries can now be generally defined by the peaks of  $\rho$  with respect to the BE or BC voltage change. The boundary of the BE SCR defines on the emitter side at the same time also the neutral emitter width

$$w_E = x \left( \frac{\partial \rho}{\partial V_{B'E}} \Big|_{V_{B'C}} = \min \right) \quad (2.23)$$

and is on the base side defined by

$$x_e = x \left( \frac{\partial \rho}{\partial V_{B'E}} \Big|_{V_{B'C}} = \max \right). \quad (2.24)$$

Similarly, the boundaries of the BC SCR are on the collector side

$$x_{C,c} = x \left( \frac{\partial \rho}{\partial V_{B'C}} \Big|_{V_{B'E}} = \min \right) \quad (2.25)$$

and on the base side

$$x_c = \min \{ x_{ci}, x_{jc} \} \quad (2.26)$$

with

$$x_{ci} = x \left( \frac{\partial \rho}{\partial V_{B'C}} \Big|_{V_{B'E}} = \max \right). \quad (2.27)$$

The limitation of  $x_c$  to the BC junction depth is necessary for separating the minority charge in the base and collector region. The above definitions are also visualized in Fig. 2.9a,b. The actual implementation of a robust regional analysis in a device simulator is somewhat more complicated due to possibly existing additional peaks from, e.g. bandgap variations, but the above equations provide the basic definitions and idea.

Another important location for the minority and depletion charge definitions is the crossover point between the carrier density variations, which is given by

$$x_m = x(\delta \rho = \delta n) = x(\delta \rho = 0). \quad (2.28)$$

It defines the so-called electrical junctions in a transistor, where the differential space charge density vanishes. Generally, this location is very close to the metallurgical junctions but it can vary with bias.

With the boundary definitions above the various region widths can now be calculated. The width of the neutral base region is

$$w_B = x_c - x_e. \quad (2.29)$$

The BE SCR width is given by

$$w_{BE} = x_e - w_E, \quad (2.30)$$

and the BC SCR width is under all bias conditions

$$w_{BC} = x_{Cc} - x_{ci}. \quad (2.31)$$

The collector injection region width is

$$w_i = \max\{x_{ci} - x_{jc}, 0\}. \quad (2.32)$$

The current dependence of the above widths is displayed in Fig. 2.10 along with  $f_T$  for reference.

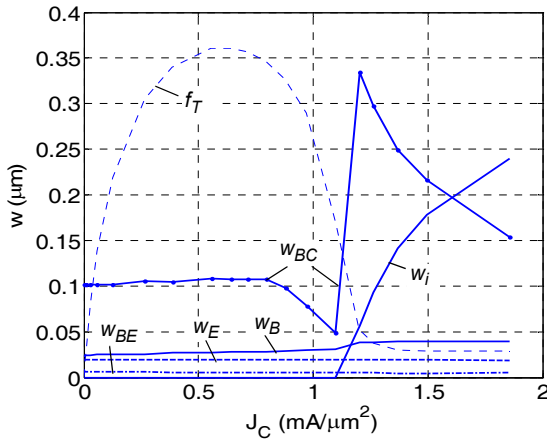


Fig. 2.10: Current dependence of the BJT region widths for  $V_{B'C'} = 0V$ . For reference,  $f_T$  has also been inserted.

Up to peak  $f_T$  all region widths change only very little, with the widest region being the base-collector SCR ( $w_{BC}$ ) and the second next the neutral base ( $w_B$ ). Beyond the peak,  $w_{BC}$  starts to collapse and then re-emerges towards the end of the collector, separated from the end of the neutral base

at  $x_{jc}$  by the injection zone  $w_i$ . Note that the injection zone forms at fairly high current densities, when  $f_T$  has dropped significantly.

Once the locations of the SCR boundaries and the resulting region widths are known, the small-signal quantities related to the neutral and space-charge regions can be easily calculated. At any given bias point, the total storage time is given from (2.19),

$$\tau_{m\Sigma} = \tau_m(L_x) = \left. \frac{dQ_m}{dI_T} \right|_{V_{CE}} = \left. \frac{dQ_m}{dV_{B'E'}} \right|_{V_{CE}} \frac{1}{g_m} = \frac{qA_E}{g_m} \int_0^{L_x} \delta m dx, \quad (2.33)$$

in which now the minority carrier charge at a CE short is inserted in order to be consistent with the calculation of  $f_T$  and the associated measurable transit time. In above equation,

$$g_m = \left. \frac{dI_T}{dV_{B'E'}} \right|_{V_{CE}} \quad (2.34)$$

is the quasi-static transconductance, and  $I_T$  is the quasi-static transfer current. Notice, that  $\tau_{mS}$  as defined is generally not equal to the transit time  $\tau_T$  used as a compact model parameter. This will be discussed later.

If the upper integration limit  $L_x$  is replaced by an arbitrary location  $x$  within the transistor, the accumulated minority transit time is obtained,

$$\tau_m(x) = qA_E \int_0^x \left. \frac{\partial m}{\partial I_T} \right|_{V_{CE}} d\xi, \quad (2.35)$$

from which the contributions of each region can be calculated with the known region boundaries. Figure 2.9c displays  $\tau_m(x)$  for the selected density values, including a very high current density at which an injection region  $w_i$  exists. In detail, the following storage time components can be distinguished. The storage time reads for the neutral emitter

$$\tau_{mE} = qA_E \int_0^{w_E} \left. \frac{\partial m}{\partial I_T} \right|_{V_{CE}} dx, \quad (2.36)$$

the BE SCR

$$\tau_{mBE} = qA_E \int_{w_E}^{x_e} \frac{\partial m}{\partial I_T} \bigg|_{V_{CE}} dx, \quad (2.37)$$

the neutral base region

$$\tau_{mB} = qA_E \int_{x_e}^{x_c} \frac{\partial m}{\partial I_T} \bigg|_{V_{CE}} dx, \quad (2.38)$$

the BC SCR

$$\tau_{mBC} = qA_E \int_{x_{ci}}^{x_{Cc}} \frac{\partial m}{\partial I_T} \bigg|_{V_{CE}} dx, \quad (2.39)$$

and the neutral collector region

$$\tau_{mC} = qA_E \int_{x_{jc}}^{L_x} \frac{\partial m}{\partial I_T} \bigg|_{V_{CE}} dx. \quad (2.40)$$

A small-signal equivalent circuit usually contains capacitances. The capacitances associated with minority charge storage follow directly from (2.19) and are defined according to (2.16) for BE and BC short, respectively:

$$C_{mE} = qA_E \int_0^{L_x} \frac{\partial m}{\partial V_{B'E}} \bigg|_{V_{B'C}} dx, \quad (2.41)$$

$$C_{mC} = qA_E \int_0^{L_x} \frac{\partial m}{\partial V_{B'C}} \bigg|_{V_{B'E}} dx. \quad (2.42)$$

The depletion capacitances are defined by (2.22). Since in a real transistor the space-charge is not confined to a depletion region, it is sometimes not easy to assign a carrier density variation to a particular junction and associated charge component. As mentioned earlier, at higher current densities the mobile charge in the BC SCR influences the BC depletion charge, although the variation is controlled by  $V_{B'E}$ . This problem of assigning a charge variation to the proper junction can be solved by using the most right expression of (2.22) instead of the middle expression. In this case, the

integration is performed always in the electrical regions of the emitter and the collector rather than in the base; i.e., for each terminal voltage variation there are in principle two depletion charge variations, one in the emitter and one in the collector. As a consequence, the BE voltage controlled depletion capacitances are given by

$$C_{jE} = \left. \frac{\partial Q_{jE}}{\partial V_{B'E}} \right|_{V_{B'C}} = qA_E \int_0^{x_{me}} \left. \frac{\partial(n-p)}{\partial V_{B'E}} \right|_{V_{B'C}} dx, \quad (2.43)$$

$$C_{cE} = \left. \frac{\partial Q_{jcE}}{\partial V_{B'E}} \right|_{V_{B'C}} = qA_E \int_{x_{mc}}^{L_x} \left. \frac{\partial(n-p)}{\partial V_{B'E}} \right|_{V_{B'C}} dx. \quad (2.44)$$

Here,  $C_{jE}$  represents the space-charge variation associated with the BE junction and, hence, corresponds to the usual BE depletion capacitance that can be obtained from a small-signal measurement directly between the B and E terminal. In contrast,  $C_{cE}$  represents the space-charge variation at the BC SCR caused both by a variation of mobile carriers within this SCR and by a variation of the voltage drop across the adjacent neutral collector region that is induced by the change of the transfer current with the change of  $V_{B'E}$ . In other words,  $C_{cE}$  contains the delay through the BC SCR and the time constant  $r_{Ci}C_{jC}$ . These two time constants are included automatically in the measured transit time  $\tau_f$  if the standard determination method is employed (e.g. [42,43]). Converting  $C_{cE}$  to a storage time,

$$\tau_{CcE} = \frac{C_{cE}}{g_m}, \quad (2.45)$$

and adding this to the BC minority time constant (2.39) leads to the modified BC storage time

$$\tau_{BC} = \tau_{mBC} + \tau_{CcE}, \quad (2.46)$$

that allows a direct comparison of  $\tau_{mS}$  from (2.33) with the measured transit time  $\tau_f$ .

In a similar way, the BC voltage controlled depletion capacitances are given by

$$C_{jC} = \left. \frac{\partial Q_{jC}}{\partial V_{B'C}} \right|_{V_{B'E}} = qA_E \int_{x_{mc}}^{L_x} \left. \frac{\partial(n-p)}{\partial V_{B'C}} \right|_{V_{B'E}} dx, \quad (2.47)$$

$$C_{eC} = \left. \frac{\partial Q_{jeC}}{\partial V_{B'C}} \right|_{V_{B'E}} = qA_E \int_0^{x_{me}} \left. \frac{\partial(n-p)}{\partial V_{B'C}} \right|_{V_{B'E}} dx. \quad (2.48)$$

Here,  $C_{jC}$  corresponds to the usual BC depletion capacitance while the cross-controlled capacitance  $C_{eC}$  represents the impact of mobile carriers on the BE depletion charge. Due to the very high doping concentration in the emitter and - for useful operating conditions - negligible feedback of a  $V_{B'C}$  change on the space-charge density around the BE junction,  $C_{eC}$  is very small and can generally be neglected even in inverse operation.

With the capacitances defined above, the hole charge variation (2.16) in the transistor can now be written as

$$dQ_p = (C_{mE} + C_{jE} + C_{eE})dV_{B'E} + (C_{mC} + C_{jC} + C_{eC})dV_{B'C} \quad (2.49)$$

and is related on one hand to measurable small-signal quantities and on the other hand to carrier densities through the capacitance definitions above. Notice that neither (2.49) nor (2.41)-(2.48) define the location of the capacitances in an equivalent circuit! Only  $C_{jE}$  and  $C_{jC}$  can be physically clearly associated with the corresponding junctions. An assignment of charge variations to equivalent circuit nodes, that is consistent with measurement (always performed at a finite frequency), requires the consideration of the time or frequency dependence of terminal currents via a solution of the continuity equation.

The storage times and capacitances defined in the regional approach above can be calculated directly from quasi-static device simulation. Figure 2.11 gives a feel for the relative importance of the intrinsic capacitances in a modern BJT. Beyond a certain critical current density the capacitance  $C_{mE}$ , representing the minority carrier storage controlled by  $V_{B'E}$ , starts to become *much larger than all other* contributions.

Figure 2.12 shows the current dependent storage times of the BJT for different voltages  $V_{B'C}$ . At low injection  $\tau_{mE}$ ,  $\tau_{mB}$  and  $\tau_{BC}$  remain basically current independent, while the BE SCR related storage time  $\tau_{BE}$  dominates the overall storage time, but drops quickly so that its influence

on the low-current *charge* is very limited. At high injection, the neutral base storage time  $\tau_{mB}$  increases - as expected - at the same rate as the neutral collector storage time  $\tau_{mC}$  due to the Kirk effect. The neutral emitter storage time  $\tau_{mE}$  increases also, but at a smaller rate due to the different small-signal current gain reduction. Another good indication of the onset of high-current effects, besides the sudden increase of  $\tau_{mC}$ , is the small peak in the modified BC storage time  $\tau_{BC}$ .

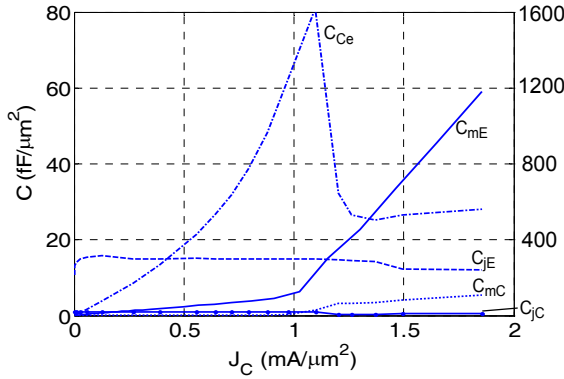


Fig. 2.11: Bias dependence of the capacitances calculated from the regional approach for the 1D BJT ( $V_{B'C'} = 0V$ ). Note that the scale between  $C_{mE}$  (right axis) and the other capacitances (left axis) differs by a factor of 20.

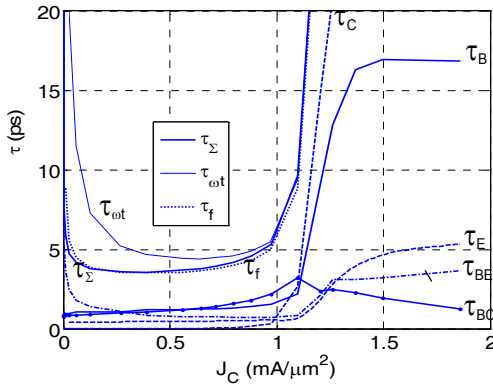


Fig. 2.12: Bias dependence of BJT storage time component for  $V_{B'C'} = 0V$ . Also, the curves for  $\tau_{m\Sigma}$ ,  $\tau_{wt} = 1/(2\pi f_T)$ ,  $\tau_f$  (from (2.53)) and are inserted for comparison.



### 2.4.2 Relation between charge storage components and measurement

The goal of the regional approach described above is to link the dynamic distribution of carrier densities to measurable small-signal terminal currents and voltages via small-signal quantities (capacitances and storage times). Such a link provides the basis for developing analytical equations that can be used in practical applications of compact models, where a relation between the stored charges and measurable quantities is required. Charges can be measured from small-signal S-parameters via capacitances and storage times. Below, important small-signal quantities are calculated on one hand from carrier densities according to the equations given earlier and on the other hand from terminal currents employing the respective standard measurement methods.

The "measured" capacitance is obtained from the frequency dependent y-parameters of the 1D transistor

$$C_{jE} = \frac{Im\{y_{11}\} + Im\{y_{12}\}}{\omega}. \quad (2.50)$$

These results agree up to a forward bias of about  $V_{B'E'} = 0.6V$  with those calculated directly from carrier densities according to the regional approach (2.43). Beyond above voltage the minority charge and associated diffusion capacitance starts to dominate so that the depletion capacitance cannot be determined anymore from (2.50). A very similar result is obtained for the BC depletion capacitance. In this case, the DC base-emitter voltage is kept at 0V according to the cold measurement method in order to eliminate any impact of mobile carriers on the BC depletion charge. The "measured" capacitance

$$C_{jC} = -\frac{Im\{y_{12}\}}{\omega} \quad (2.51)$$

agrees very well with the one calculated from carrier densities according to (2.47) until beyond a certain forward bias the mobile carrier density starts to become significant in the BC SCR.

The transit frequency can be calculated accurately and very efficiently from y-parameters according to [41]

$$f_T = \frac{f}{\text{Im}\{y_{21}/y_{21}\}}, \quad (2.52)$$

where  $f$  is the measurement frequency and  $\beta = y_{21}/y_{11}$  is the common emitter small-signal current gain. Above equation, a.k.a. spot-frequency method, allows to determine  $f_T$  at a single frequency without the need to fit the data to a low-pass behavior and is particularly attractive also for device and circuit simulation. Using charge-control theory  $f_T$  can be related directly to the total hole charge variation  $dQ_p$  in (2.20) according to (e.g. [44])

$$\frac{1}{2\pi f_T} = \left. \frac{dQ_p}{dI_C} \right|_{V_{CE}}. \quad (2.53)$$

The validity of calculating the small-signal quantities from quasi-static carrier distributions is confirmed by the comparison in Fig. 2.13. The corresponding curves are on top of each other over the entire operating range of interest. This is also consistent with the separate comparisons of depletion capacitances as well as the overall storage and transit time.

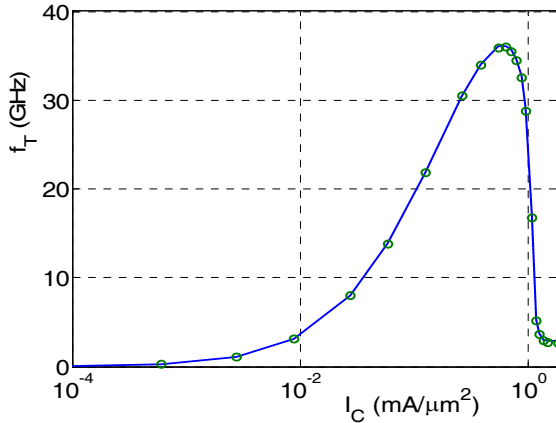


Fig. 2.13: Current dependence of the transit frequency for the 1D BJT: comparison between frequency dependent measurement method (symbols, eq. (2.52)) and quasi-static calculation (solid line, eq. (2.53));  $V_{B \cdot C}/V = 0V$ .

Finally, the measured transit time is typically obtained from the standard determination method (e.g. [42, 43]) according to

$$\tau_f = \frac{1}{2\pi f_T} - \frac{\sum C_v}{g_m}, \quad (2.54)$$

where  $\sum C_n$  in general represents all capacitances connected to the base terminal. A comparison of  $\tau_f$  with the total storage time  $\tau_{ms}$ , calculated from the regional contributions of quasi-static carrier distributions, shows good agreement in Fig. 2.12. As a consequence of this and the other results presented in this chapter, measured charge-storage elements can be linked directly to carrier distributions via the regional analysis. This facilitates the development of physics-based model equations.

## 2.5 References

- [1] J. Cressler (ed.), "Silicon heterostructure Handbook", CRC Press NY, 2005.
- [2] C.-T. Sah, "Fundamentals of Solid-State Electronics", World Scientific, Singapore, 1991.
- [3] T.H. Ning, "History and future perspective of the modern silicon bipolar transistor", IEEE Trans. Electron Dev. Vol. 48, pp. 2485-2491, 2001.
- [4] C.-T. Sah, "Evolution of the MOS Transistor - From Conception to VLSI", Proc. IEEE, Vol. 76, No. 10, pp. 1280-1326, 1988.
- [5] A.R. Alvarez, "Future Trends In BiCMOS Technology", Microelectronic Engineering, pp. 493-500, 1991.
- [6] J. Matthews and A. Blakeslee, "Defects in epitaxial multilayers", Journal of Crystal Growth, Vol.27, pp. 118-125, 1974, and Vol. 29, pp. 273-280, 1975.
- [7] B. Orner et al., "A 0.13mm BiCMOS technology featzuring a 200/280 GHz ( $f_T/f_{max}$ ) SiGe HBT", IEEE Bipolar/BiCMOS Circuits and Technology Meeting, pp. 203-206, 2003.
- [8] J. Böck, H. Schäfer, K. Aufinger, R. Stengl, S. Boguth, R. Schreiter, M. Rest, H. Knapp, M. Wurzer, W. Perndl, T. Bottner, and T.Meister, "SiGe bipolar technology for automotive radar applications", IEEE Bipolar/BiCMOS Circuits and Technology Meeting, pp. 84-87, 2004.
- [9] P. Chevalier, C. Fellous, L. Rubaldo, F. Pourchon, S. Pruvost, R. Beerkens, F. Saguin, N. Zerounian, B. Barbalat, S. Lepilliet, D.

- Dutartre, D. Céli, I. Telliez, D. Gloria, F. Aniel, F. Danneville and A. Chantre, “230-GHz Self-Aligned SiGeC HBT for Optical and Millimeter-Wave Applications”, in *IEEE Journal of Solid-State Circuits*, Vol. 40, 2005, pp. 2025–2034.
- [10] M. Racanelli, P. Kempf, “SiGe BiCMOS technology for RF circuit applications”, *IEEE Trans. Electron Dev.* Vol. 52, pp. 1259-1270, 2005.
- [11] M. Khater, J.S. Rieh, T. Adam, A. Chinthakindi, J. Johnson, R. Krishnasamy, M. Meghelli, F. Pagette, D. Sanderson, C. Schnabel, K.T. Schonenberg, P. Smith, K. Stein, A. Stricker, S.J. Jeng, D. Ahlgren and G. Freeman, “SiGe HBT Technology with  $f_{\text{max}}/f_T=350/300$  GHz and Gate Delay Below 3.3 ps”, *IEDM Tech. Dig.*, 2004, pp. 247–250.
- [12] P. Chevalier et al., “300 GHz  $f_{\text{max}}$  self-aligned SiGeC HBT optimized towards CMOS compatibility”, *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 120–123, Oct. 2005.
- [13] A. Fox, B. Heinemann, R. Barth, B. Bolze, J. Frews, U. Haak, D. Knoll, B. Kuck, R. Kurps, A. Marcsh Meyer, H. Richter, H. Rucker, P. Schley, D. Schmidt, B. Tillack, G. Weidner, C. Wipf, D. Wolansky, Y. Yamamoto, “SiGe HBT module with 2.5ps gate delay”, *IEDM Digest*, pp. 731-734, 2008.
- [14] R. Krithivasan, Y. Lu, J. D. Cressler, J.-S. Rieh, M. H. Khater, D. Ahlgren, G. Freeman, “Half-Terahertz Operation of SiGe HBTs,” *IEEE Electron Device Lett.*, vol. 27, no. 7, pp. 567-569, 2006.
- [15] N. Zerounian, E. Garcia, F. Aniel, P. Chevalier, B. Geynet, and A. Chantre, “SiGe HBT featuring  $f_T > 600\text{GHz}$  at cryogenic temperature”, *ECS Transactions*, Vol. 16 (10), pp. 1069-1077, 2008.
- [16] W. Shockley, “Circuit element utilizing semiconductor material”, *US Patent 2569347*, Sept. 1951.
- [17] H. Krömer, “Heterostructure bipolar transistors and integrated circuits”, *IEEE Proc.*, Vol. 70, pp. 13-25, 1982.
- [18] W. Dumke, J. Woodall, V. Rideout, “GaAs-(GaAl)As heterojunction transistor for high-frequency operation”, *Solid-State Electronics*, Vol. 15, pp. 1339-1343, 1972.
- [19] M.F. Chang, P. Asbeck, K. Wang, G. Sullivan, N. Sheng, J.A. Higgins, D. Miller, “AlGaAs/GaAs heterojunction bipolar transistors fabricated using a self-aligned dual-lift-off process”, *IEEE Electron Dev. Lett.*, Vol. 8, pp. 860-862, 1987.

- [20] W Hafez, W. Snodgrass, M. Feng, “12.5 nm base pseudomorphic heterojunction bipolar transistors achieving  $f_T=710\text{GHz}$  and  $f_{max}=340\text{GHz}$ ”, *Appl. Phys. Lett.*, Vol. 87, 252109, 2005.
- [21] H. Liu, O. Ostinelli, Y. Zheng, C. Bolognesi, “600 GHz InP/GaAsSb/InP DHBTs grown by MOCVD with a Ga(As,Sb) graded base and  $f_T \times BV_{CEO} > 2.5 \text{ THz-V}$  at room temperature”, *IEEE IEDM Tech. Digest*, pp. 667-670, 2007.
- [22] M. Rodwell, E. Lobisser, M. Wistey, V. Jain, A. Baraskar, E. Lind, J. Koo, Z. Griffith, J. Hacker, M. Urtega, D. Mensa, R. Pierson, B. Brar, “THz bipolar transistor circuits: Technical feasibility, technology development, integrated circuit results”, *Proc. IEEE CSICS*, pp. 3-5, 2008.
- [23] M. Lundstrom, “Fundamentals of carrier transport”, 2nd ed., Cambridge, 2000;  
see also first ed.: Addison-Wesley, Modular Series on Solid-State Devices, Vol. X, 1990.
- [24] C. Jacoboni, P. Lugli, “The Monte-Carlo method for semiconductor device simulation”, Springer, 1989.
- [25] C. Jungemann, “A Deterministic Approach to RF Noise in Silicon Devices Based on the Langevin–Boltzmann Equation”, *IEEE Trans. Electron Dev.*, Vol. 54, pp. 1185-1192, 2007.
- [26] T. Grassner et al., “A review of hydrodynamic and energy transport models for semiconductor device simulation”, *Proc. IEEE*, Vol. 91, No.2, pp. 251-271, 2003.
- [27] C. Jungemann, B. Meinerzhagen, “Hierarchical Device Simulation: The Monte Carlo Perspective”, Springer New-York, 2003.
- [28] S. Selberherr, “Analysis and simulation of semiconductor devices”, Springer, 1981.
- [29] M. Kurata, “Numerical analysis of semiconductor devices”, Lexington Press, 1982.
- [30] M.S. Lundstrom and R. Schuelke, “Modeling semiconductor heterojunctions in equilibrium”, *Solid-State Electronics*, Vol. 25, pp. 683-691, 1982.
- [31] M. Schröter, “Transient and small-signal high-frequency simulation of numerical device models embedded in an external circuit”, *COMPEL*, Vol. 10, No. 4, pp. 377-378, 1991.

- [32] M. Schröter, S. Lehmann, D. Celi, "Non-standard geometry scaling effects in high-frequency SiGe bipolar transistors", Proc. WCM, International NanoTech Meeting, Santa Clara, pp. 603-608, 2007.
- [33] Dessim Manual, Sentaurus Device User Guide V. Z-2007.03, Synopsys, 2007.
- [34] C. Jungemann, "MONJU User's Guide V. 1.6.5", ITEM, TU Braunschweig, 2006.
- [35] G. Sasso, internal report, CEDIC, 2008.
- [36] G. Wedel, internal report, CEDIC, 2008.
- [37] C. Jungemann, B. Neinhüs, B. Meinerzhagen, "Comparative study of electron transit times evaluated by DD, HD, and MC device simulation for a SiGe HBT", IEEE Trans. Electron Devices, Vol. 48, No. 10, pp. 2216-2220, 2001.
- [38] J.J. van den Biesen, "A Simple Regional Analysis of Transit Times in Bipolar Transistors", Solid-State Electronics, Vol. 29, pp. 529-534, 1986.
- [39] D.J. Roulston, S.G. Chamberlain and J. Sehgal, "Simplified Computer-Aided Analysis of Double-Diffused Transistors Including Two-Dimensional High-Level Effects", IEEE Trans. Electron Dev., Vol. ED-19, pp. 809-820, 1972.
- [40] M. Schröter, H. Tran, "Charge-storage related parameter calculations for Si and SiGe bipolar transistors from device simulation", Proc. WCM, International NanoTech Meeting, Boston (MA), pp. 735-740, 2006.
- [41] H.K. Gummel, "On the definition of the cutoff frequency  $f_T$ ", Proc. IEEE, Vol. 57, pp. 2159, 1969.
- [42] H.-M. Rein, "Proper Choice of the Measuring Frequency for Determining  $f_T$  of Bipolar Transistors", Solid-State Electronics, Vol. 26, pp. 75-83 and pp. 929, 1983.
- [43] M. Schröter und H.-M. Rein "Transit Time of High-Speed Bipolar Transistors in Dependence on Operating Point, Technological Parameters, and Temperature", Proc. IEEE Bipolar Circuits and Technology Meeting, Minneapolis, pp. 250-253, 1989.
- [44] M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors with special regard to high current densities and two-dimensional effects", Ph.D. thesis (in German), Ruhr-University Bochum, Bochum, Germany, 1988.

## **Chapter 3**

# **Theory of Homojunction Bipolar Transistors**

### 3.1 Operation principle

The discussion below focuses on the forward operation mode, which by far is most relevant for practical applications. The most simple (classical) transistor theory (e.g. [1-3,4]) predicts just an ideal exponential increase of the collector current  $I_C$  with the controlling voltage  $V_{B'E'}$ , as well as a constant transit time  $\tau_{f0}$  over the entire voltage and current range. As shown in Fig. 3.1, which summarizes schematically the 1D device simulation results of the previous chapter for the collector current and transit time, classical theory becomes inaccurate at medium and high current densities. Especially the medium current range is important for various high-speed circuits.

In the literature, there is no exact definition of the boundaries between low, medium and high current densities. In this book, the beginning of the medium current region is defined by the current dependent increase of the transit time  $\tau_f$ . Furthermore, the boundary to the high current region is defined by a critical current  $I_{CK}$  (derivation see later) which indicates the strong increase of  $\tau_f$ .

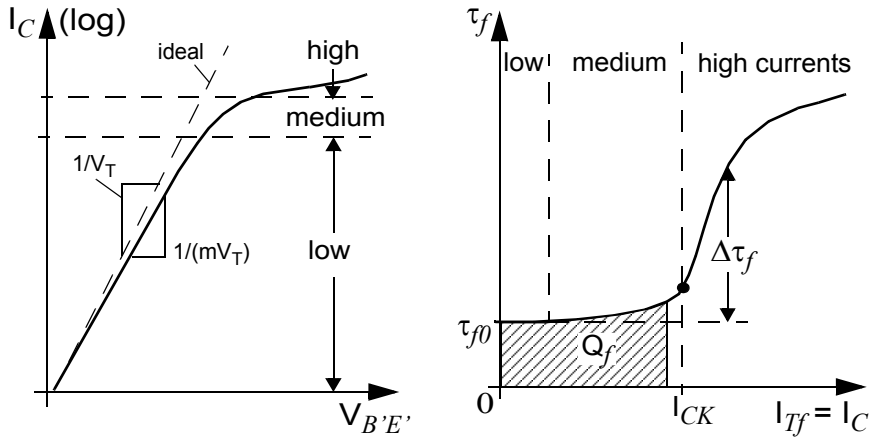


Fig. 3.1: Definition of the bias regions and important transistor variables: (a) DC collector (transfer) current vs. BE voltage;  $m \approx 1$  is a non-ideality coefficient, which can include the impact of recombination effects. Note though that these are negligible in modern Si-based BJTs and HBTs. (b) minority transit time and associated charge as a function of current.  $V_{B'C'}$  or  $V_{C'E'} = \text{const.}$



Prior to any theoretical treatment, first a qualitative analysis shall be performed by means of 1D device simulation results. This analysis provides a better feeling for the different operating regions and allows to derive suitable analytical approximations. For the discussions and derivations in this book it is useful to define the (internal) collector voltage

$$V_{ci} = -\int_{x(N_{B \max})}^{L_x} E_x dx = V_{DCi} - V_{B'C} \approx V_{CE} - V_{CEs}. \quad (3.1)$$

The right most expression results from inserting  $V_{B'C} = V_{B'E} - V_{CE}$ , then approximating  $V_{B'E}$  by the BE junction built-in voltage  $V_{DEi}$ , and finally defining  $V_{CEs} = V_{DEi} - V_{DCi}$  as the internal CE saturation voltage.

For explaining the observed characteristics sketched in Fig. 3.1 it is useful to consider first the case of "high" collector voltages. Afterwards, the case of "low" voltages, which is a little more complicated, is discussed. From this qualitative description of the occurring effects the critical current can be directly derived. It is also convenient for the considerations to keep the voltage  $V_{ci}$  at a constant value and to vary only the current density  $I_T/A_E = I_C/A_E$ . Again, since the 1D case is discussed the designation current is used synonymously with current density.

### 3.1.1 "High" collector voltage

Figure 3.2(b),(c) shows the behavior of electric field and electron density vs. location in the base-collector region for the case of "high" voltages, which will be more exactly defined later. In this book, the variable  $E_{jc} = -E_x(x_{jc})$  is used for the electric field at the BC junction. At low current densities, the electric field has its peak at the junction and already consumes most or all of the epitaxial collector region (curve 1). With increasing current density the electron density in the BC SCR increases according to

$$n_c = \frac{I_T}{qA_E v_c}. \quad (3.2)$$

According to Fig. 3.2a, the assumption of pure drift transport is justified and allows to substitute  $v_c$  by the saturation velocity  $v_{sn}$ . Inserting this into Poisson's equation gives

$$\frac{dE_x}{dx} = \frac{qN_{Ci}}{\epsilon} \left( 1 - \frac{I_T}{I_{lim}} \right) \quad (3.3)$$

with the characteristic current

$$I_{lim} = qA_E N_{Ci} v_{sn}. \quad (3.4)$$

In the low-current region the condition  $I_T \ll I_{lim}$  holds.

Once  $I_T$  is no longer negligible compared to  $I_{lim}$  the electron density starts to partially compensate the (fixed positively charged) donor density  $N_{Ci}$ . As a result, with increasing current the slope of the field starts to decrease leading also to a decrease of the peak field and, thus,  $E_{jc}$  as can be observed in curve 2 of Fig. 3.2b). In this case, the SCR edge on the collector side has just reached the buried layer; i.e. the SCR has expanded and now consumes the entire collector region. If  $I_T$  is further increased the slope of  $E_x$  passes through zero (curve 3) once  $n$  reaches  $N_{Ci}$ , i.e. has fully compensated the (fixed) donor density. However, despite a zero slope and the drop of  $E_{jc}$  the field is still large enough to move the carriers with saturation velocity (cf. Fig. 3.2a) throughout the entire collector region.

Further increase of the current density leads, according to (3.3), to a negative slope (curve 4) and a continuing decrease of  $E_{jc}$  until the field approaches

$$E_{lim} = v_{sn} / \mu_{nCi0} \quad (3.5)$$

with  $\mu_{nCi0}$  as doping dependent low-field mobility in the collector. At  $E_{lim}$  the velocity  $v_c$  starts to become smaller than  $v_{sn}$  (cf. Fig. 3.2a). At this point, the limit of the medium current density range is reached and the transfer current cannot be carried any more by drift alone. For a given current density the lower drift velocity needs to be compensated by a larger electron density than the one needed for  $v_c = v_{sn}$ . In other words, the electron density at  $x_{jc}$  now starts to increase more than proportional to  $I_T$  (curve 5). This in turn leads to a more than proportional decrease of  $E_{jc}$  and increase in charge with increasing current density.

From the field distribution follows that most of the collector voltage drop now occurs towards the end of the collector region rather than around the junction (as in curves 1 and 2). The resulting small voltage left across the junction region leads to a potential distribution that is equivalent to that

in a forward biased pn-junction. As a consequence, hole injection starts with the hole density also increasing rapidly with  $I_T$ .

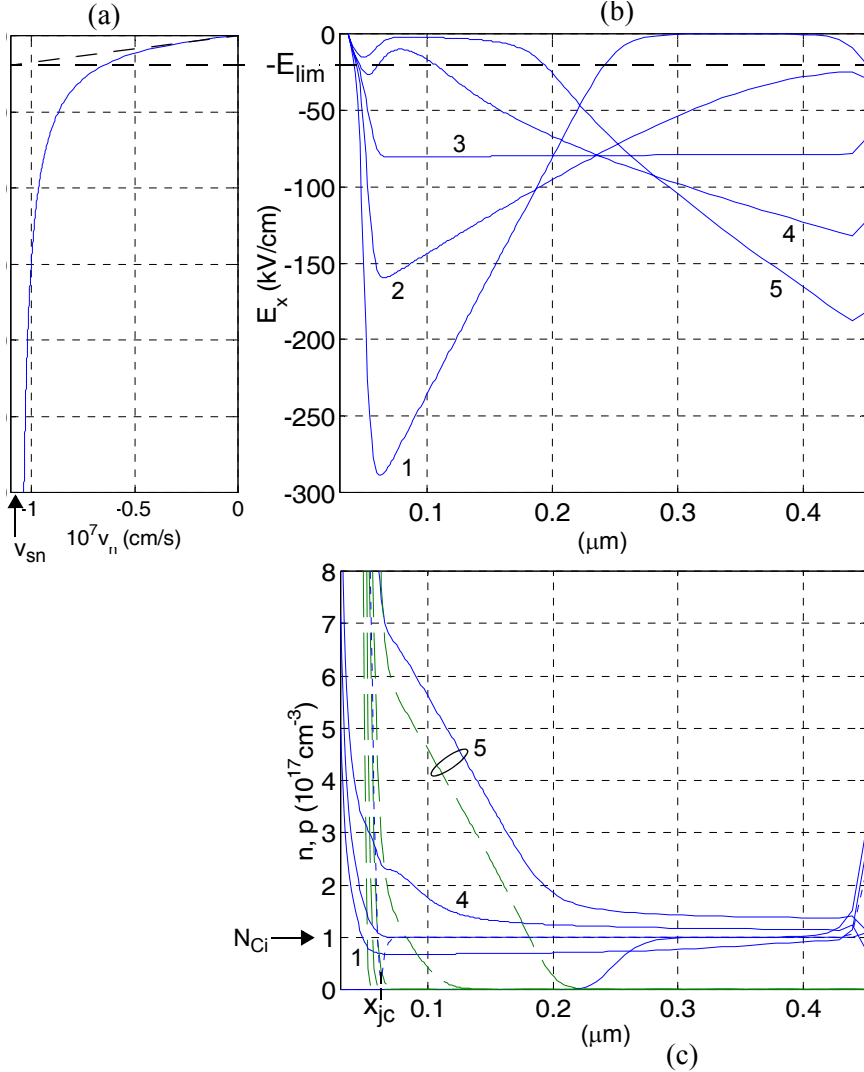


Fig. 3.2: Results of 1D BJT device simulation within the base-collector region in case of a "high" voltage  $V_{ci}$ : (a) electron drift velocity  $v_n$ ; (b) electric field; (c) electron density (solid lines), hole density (dashed lines) and doping (dotted line). The curves 1...5 represent different currents  $I_C = (0.72 \cdot 10^{-3}, 1.08, 1.5, 1.8, 2.2) \text{ mA}$ ;  $V_{B'C'} = -2.3 \text{ V}$  and  $E_{lim} = 15 \text{ kV/cm}$ .

Finally, at very high current densities both carrier densities become much larger than  $N_{Ci}$  (curve 5) leading to a neutral region around the junction - the so-called injection region - the boundary of which moves towards the buried layer with current density. The injection of holes into this region leads to a dramatic reduction in transistor operating speed through the excess minority charge storage and the resulting decrease of the transconductance.

The boundary between the medium and high-current region (curve 4) defines the critical current  $I_{CKh}$ , which can be derived from (3.3). Be  $E_{wc} = -E_x(w_{Ci})$  the field at the buried layer end, then (3.3) reads

$$\frac{dE_x}{dx} \cong -\frac{E_{wc} - E_{lim}}{w_{Ci}} = \frac{qN_{Ci}}{\epsilon} \left(1 - \frac{I_{CKh}}{I_{lim}}\right). \quad (3.6)$$

Inserting  $E_{wc}$  from the condition

$$V_{ci} = -\int_0^{w_{Ci}} E_x dx' \cong \frac{E_{lim} + E_{wc}}{2} w_{Ci} \quad (3.7)$$

and solving for the critical current yields [6]

$$I_{CKh} = I_{lim} \left(1 + \frac{V_{ci} - V_{lim}}{V_{PT}}\right) \quad (3.8)$$

with the collector punch-through voltage

$$V_{PT} = \frac{qN_{Ci}}{2\epsilon} w_{Ci}^2. \quad (3.9)$$

(3.8) leads, in contrast to the literature (e.g. [7, 8, 9]), to much more realistic values of  $I_{CKh}$  since the finite field at the boundary to the high-current region is taken into account by the voltage

$$V_{lim} = E_{lim} w_{Ci}. \quad (3.10)$$

In Fig. 3.3 internal variables, which are of interest for the derivation of compact model equations, are displayed within the entire 1D transistor region and for the same bias points as listed in Fig. 3.2. Figure (a) contains the vertical component of the electron current density,  $J_{nx}$ . Horizontal lines indicate that the impact of recombination is negligible. Hence, only for the bias point far in the high-current region a small increase of  $J_{nx}$  in the neu-

tral emitter region is visible (caused by Auger recombination). Figure (b) exhibits the quasi-Fermi potential for electrons and holes. As can be seen clearly,  $\phi_p$  is constant across the entire base and emitter region (except very close to the contact) but changes in the collector depending on the current density (i.e. injection level). In contrast,  $\phi_n$  is only constant ( $= 0$ , since the E is the reference electrode) in the neutral emitter and then starts to increase at the beginning of the neutral base towards the collector. In the collector,  $\phi_n$  also depends strongly on the current density, and its decreasing slope at the BC junction reflects the decreasing electric field there. It is clearly visible that the difference  $\phi_p - \phi_n$  at the BC junction does not equal the applied voltage  $V_{B'C'}$ , but is strongly current dependent.

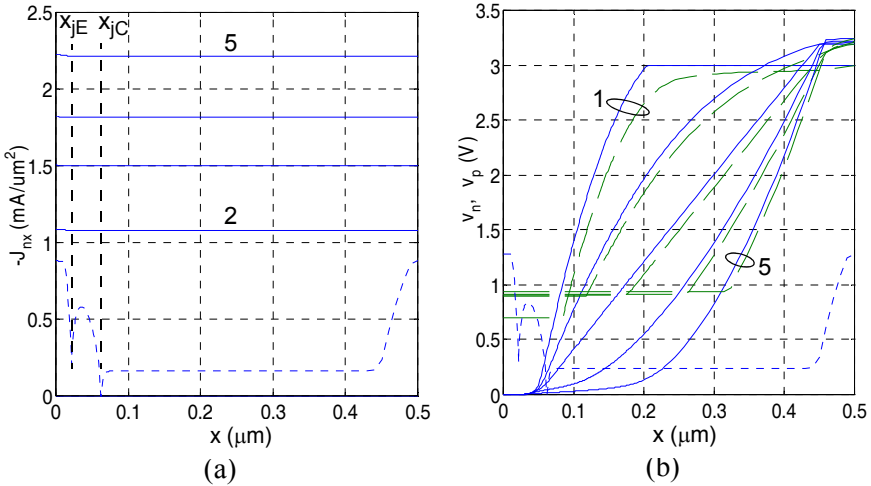


Fig. 3.3: Important internal variables of the BJT for the high-voltage case: (a) vertical electron current density. (b) Quasi-Fermi potential of electrons (solid lines) and holes (dashed lines). As reference, also the doping profile is shown (dotted lines). For the specification of bias points see Fig. 3.2.

### 3.1.2 "Low" collector voltage

At first glance, the field and carrier density distribution at "low" collector voltages in Fig. 3.4b,c looks similar to that at high voltages. Curve 1 again corresponds to an operating point at (very) low current densities with a partially depleted collector. As discussed in the previous section, a cur-

rent density increase leads to a compensation of donors and to a decrease of the field once the electron density rises up to the order of the doping density. However, the smaller starting field also leaves a larger portion of the collector undepleted which acts as an ohmic resistance. This is clearly visible in curve 2 by the non-zero electric field in the undepleted collector region. Therefore, at low voltages the increasing current density leads not only to a compensation of the space charge but also to an additional significant voltage drop in the ohmic region, which in turn reduces the voltage and field across the junction region even faster with current than in the high-voltage case. It also becomes evident that the width of the SCR decreases with current.

As for the high-voltage case, here also exists a current density that causes the electric field to be horizontal throughout (almost) the entire collector region (curve 3); this corresponds again to  $n = N_{Ci}$ . However, the value of this horizontal field, which is simply given by

$$E_{CKI} = \frac{V_{ci}}{w_{Ci}}, \quad (3.11)$$

is located within the linear (ohmic) portion of the velocity field curve as the projection to the corresponding curve in Fig. 3.4a shows. This also means that the SCR actually has disappeared. The electron current density at this point still consists mostly of a drift component and is given by [6]

$$I_{CKI} = qA_E N_{Ci} \mu_{nCi}(N_{Ci}, E_{CKI}) E_{CKI}, \quad (3.12)$$

in which a suitable mobility model needs to be inserted.

Increasing the current density beyond this level leads again to a reversal of the slope of the electric field (curve 4). Also, the field close to the junction collapses to a very small value. Due to the associated low drift velocity the electron density needs to increase more than proportional with the current density, resulting in an additional diffusion component. Furthermore, the disappearing potential barrier for holes leads to an increase of the hole density within an injection zone. The corresponding excess minority charge storage then causes a significant reduction of the transistor speed.

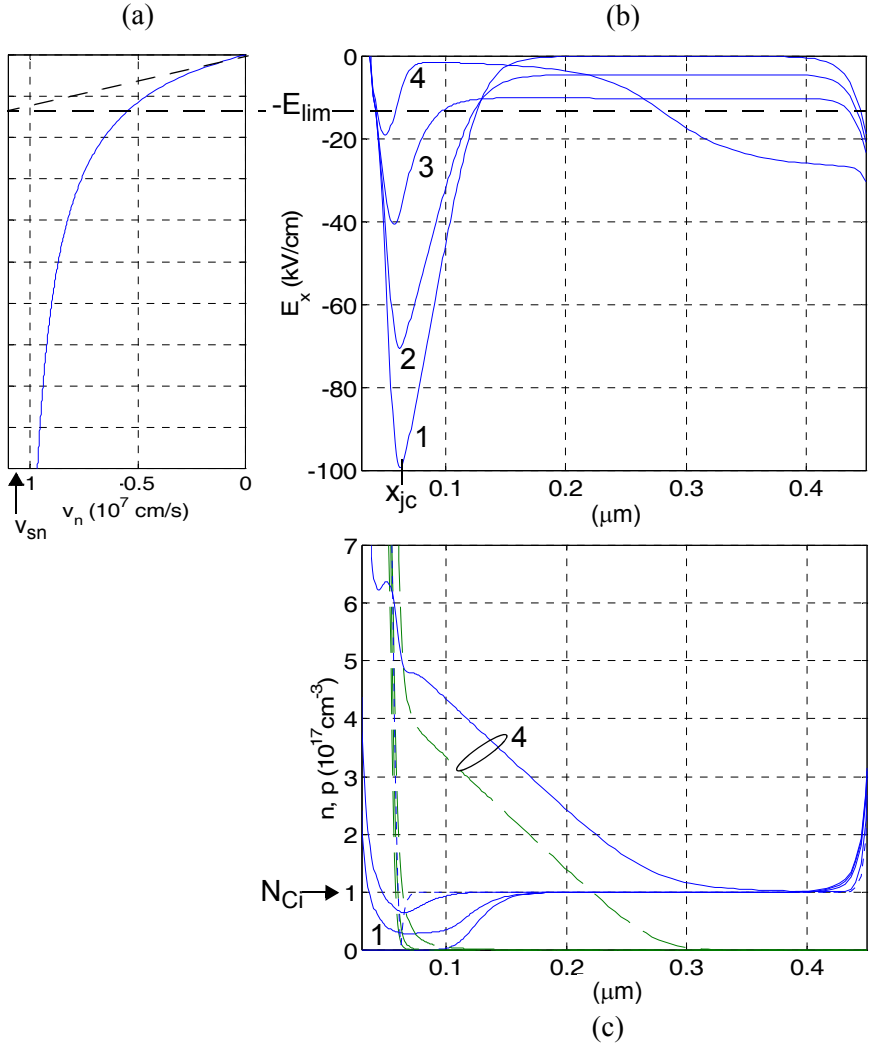


Fig. 3.4: Results of 1D BJT device simulation within the base-collector region in case of a "low" voltage  $V_{Ci}$ : (a) electron drift velocity  $v_n$ ; (b) electric field; (c) electron density (solid lines), hole density (dashed lines) and doping (dotted line). The curves 1...4 are for different current densities  $I_C = (0.64 \cdot 10^{-3}, 0.41, 0.75, 1.18)$  mA;  $V_{B'C'} = 0.4$  V. Note, that the residual field in the base is caused by the built-in field from doping and effective intrinsic concentration.

For the same bias points as listed in Fig. 3.4 relevant internal variables are displayed in Fig. 3.5 within the entire 1D transistor region. The horizontal lines of the vertical electron current density  $J_{nx}$  in Fig. 3.4a indicate a negligible impact of recombination, except in the neutral emitter for the bias point far in the high-current region. Figure 3.4b exhibits the quasi-Fermi potential for electrons and holes. In this case of forward BC voltage,  $\phi_p$  is constant across almost the entire transistor region. In contrast,  $\phi_n$  is only constant ( $= 0$ ) in the neutral emitter and then starts to increase at about  $x_e$  towards the collector. Again,  $\phi_n$  depends strongly on the current density with its decreasing slope at the BC junction reflecting the decreasing electric field there, and the difference  $\phi_p - \phi_n$  at the BC junction does not equal the applied voltage.

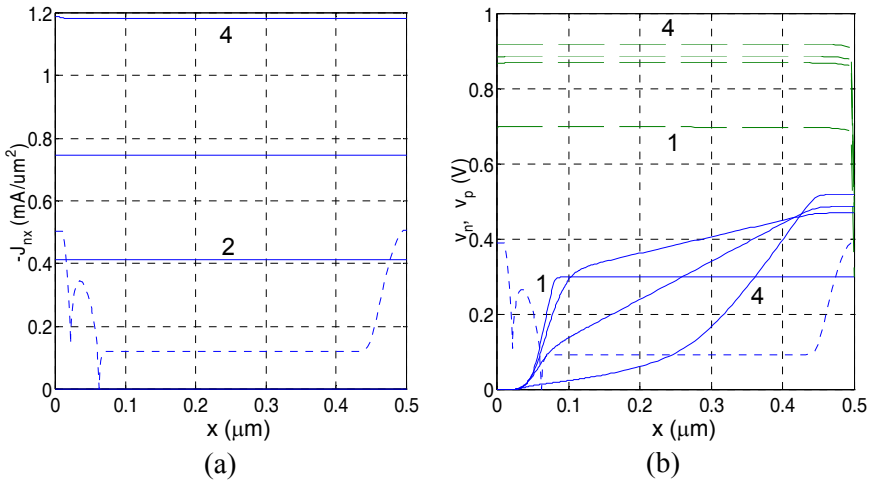


Fig. 3.5: Important internal variables of the BJT for the low-voltage case: (a) vertical electron current density. (b) Quasi-Fermi potential of electrons (solid lines) and holes (dashed lines). As reference, also the doping profile is shown (dotted lines). For the specification of bias points see Fig. 3.4.

### 3.1.3 Summary

At medium and high current densities, the electric field distribution in the collector BC SCR is influenced by the current since the associated mobile carrier density compensates the fixed doping charge density. In addi-



tion, at low voltages an ohmic voltage drop occurs in the undepleted collector portion. The SCR itself collapses at low voltages and expands at high voltages. As result of the reduction of  $E_{jc}$  at medium current densities, the width of the neutral base increases, leading to a decrease in current gain and an increase of the associated storage times in base and emitter. Once the field has dropped so much that the BC SCR loses its function of rapidly moving the minority carriers out of the base, the onset of the high current region is reached, which is defined by the critical current. In this region, minorities are injected into the collector while the field starts to build up at the buried layer. Note though that in contrast to an often found perception (e.g. [5]) homojunction transistors can be operated without significant impact on their characteristics at and beyond  $I_{lim}$  as long as  $E_{jc}$  remains 2...3 times above  $E_{lim}$ .

In the next section, the electric field in the collector is analyzed in order to obtain a better feeling for the effects at high-current densities.

### 3.2 Electric field calculations

As evident from the previous discussion classical transistor theory with focus on the (neutral) base region only fails to describe the electrical behavior of actual transistor structures. This is due to the collector region with its lower doping concentration (than in the base), which leads to a reduction of the electric field at the BC junction at high current densities. In other words, the collector region limits the current carrying capability of a transistor. This fact was realized in the early sixties (e.g. [10,11]). The most widely referenced paper on this topic is that of Kirk [10], which led to the designation *Kirk-effect* for the behavior described in Figs. 3.2 and 3.4. Since the electric field at the BC junction determines the transistor characteristics at medium and high current densities, the most important basic relations are derived in the subsequent sections.

Closed-form analytical solutions for the electric field at the BC junction,  $E_{jc}$ , and at the transition point to the buried layer,  $E_{wc}$ , can be obtained as a function of bias  $V_{B'C'}$  and  $I_T$  for different operating regions under the following assumptions:

- spatially independent collector doping concentration  $N_{Ci}$  within the collector region of width  $w_{Ci}$ ;

- the carrier velocity  $v_c$  at the BC junction equals the saturation drift velocity  $v_{sn}$ ;
- normal transistor operation with  $V_{C'E'} > V_{CEs}$ .

Below, electric field expressions are derived as a function of bias for the two fundamentally different cases of low and high collector voltage.

### 3.2.1 Fully depleted collector region - high voltage case

Under the above assumptions, Poisson's equation within the BC SCR is given by (3.3). The sketch of a realistic spatial field distribution in the collector region (cf. Fig. 3.2) is depicted in Fig. 3.6 for a sufficiently high collector voltage when punch-through occurs. For analytical treatment, the x-axis has been shifted and starts at the BC junction.

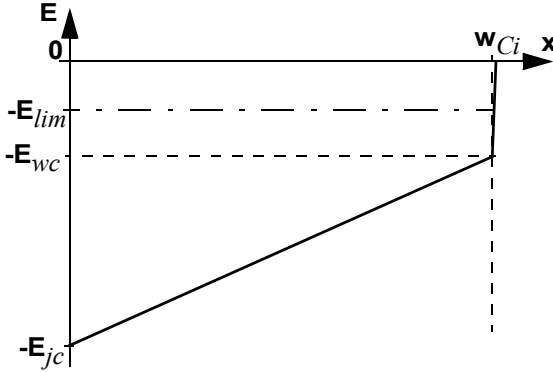


Fig. 3.6: Sketch of the spatial field distribution for a fully depleted collector at a low current density.

The relation between field and applied voltage is given by integrating Poisson's equation (3.3) twice. The first integration yields

$$\int_0^x \frac{dE}{dx} dx = \int_0^x \frac{qN_{Ci}}{\epsilon} \left(1 - \frac{I_T}{I_{lim}}\right) dx = \frac{qN_{Ci}}{\epsilon} \left(1 - \frac{I_T}{I_{lim}}\right) x,$$

which gives the spatially dependent electric field

$$E(x) = E(0) + \frac{qN_{Ci}}{\epsilon} \left(1 - \frac{I_T}{I_{lim}}\right) x. \quad (3.13)$$

Using  $E(0) = -E_{jc}$  and integrating again over the entire epi region yields

$$\int_0^{w_{Ci}} E dx = -E_{jc} w_{Ci} + V_{PT} \left( 1 - \frac{I_T}{I_{lim}} \right) = -V_{ci},$$

with the punch-through voltage  $V_{PT}$  from (3.9). The solution reads

$$E_{jc} = \frac{V_{ci} + V_{PT} \left( 1 - \frac{I_T}{I_{lim}} \right)}{w_{Ci}}. \quad (3.14)$$

The voltage dependence at negligible currents is given by

$$E_{jc0} = \frac{V_{ci} + V_{PT}}{w_{Ci}} = E_{jc00} \left( 1 + \frac{V_{B'C}}{V_{DC} + V_{PT}} \right) \quad (3.15)$$

with  $E_{jc00} = E_{jc}(V_{B'C} = 0, I_T = 0)$ . Above equation is valid for any bias point as long as (i) the punch-through condition (cf. also (3.31))

$$V_{ci} + E_{wc} w_{Ci} \geq V_{PT} \left( 1 - \frac{I_T}{I_{lim}} \right) \quad (3.16)$$

is met, and (ii) the value of  $E_{jc}$  is equal or larger than  $E_{lim}$ .

### 3.2.2 Partially depleted collector: low voltage case

The sketch of a realistic spatial field distribution in the collector region (cf. Fig. 3.4) is depicted in Fig. 3.7 as the solid line. At low current densities, the BC depletion region is still located close to the junction, and its width extends to the location  $w_{BC,c}$ , which is smaller than the epi width  $w_{Ci}$ . The remaining region with the width  $(w_{Ci} - w_{BC,c})$  is undepleted and behaves electrically like an ohmic region. It is assumed that the field within the depletion region around is large enough to cause the carriers to move with saturation drift velocity. This assumption is obviously not fulfilled in the transition region  $w_{BC,c}$ , i.e. between space charge and ohmic region. The corresponding *consistent* field curve (cf. long dashes) is also inserted and causes the field distribution to be spatially discontinuous. The short dashes represent an idealized distribution used for analytical treatment.

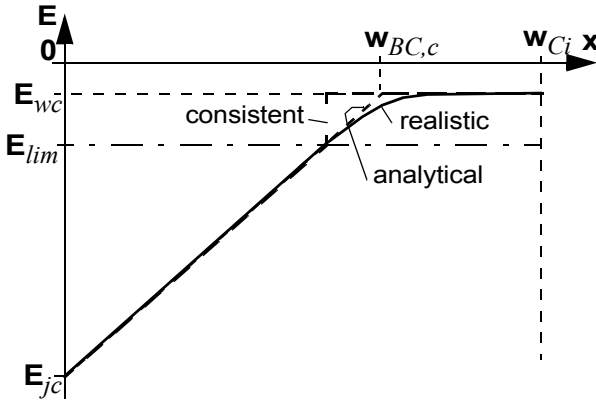


Fig. 3.7: Spatial field distribution for a partially depleted collector at a low current density.

In the ohmic region  $x \in [w_{BC,c}, w_{Ci}]$ , a linear potential drop exists,

$$V(x) - V(w_{BC,c}) = -\rho_{Ci} \frac{x - w_{BC,c}}{A_E} I_T. \quad (3.17)$$

Thus, the field is spatially constant and given by

$$E_{wc} = \rho_{Ci} \frac{I_T}{A_E} = \frac{I_T}{A_E q \mu_{nCi}(E_{wc}) N_{Ci}} \quad (3.18)$$

with the often used field dependent electron mobility expression,

$$\mu_{nCi}(E_{wc}) = \frac{\mu_{nCi0}}{\sqrt{1 + \left(\frac{(-E_{wc})}{E_{lim}}\right)^2}}. \quad (3.19)$$

Inserting into (3.18) yields

$$E_{wc} = \frac{I_T}{A_E q \mu_{nCi0} N_{Ci}} \sqrt{1 + \left(\frac{E_{wc}}{E_{lim}}\right)^2},$$

which results in a quadratic equation for  $E_{wc}$ ,

$$E_{wc}^2 = \left(\frac{I_T}{A_E q \mu_{nCi0} N_{Ci}}\right)^2 \left[1 + \left(\frac{E_{wc}}{E_{lim}}\right)^2\right] = \left(\frac{I_T}{I_{lim}} E_{lim}\right)^2 \left[1 + \frac{E_{wc}^2}{E_{lim}^2}\right],$$

with the explicit solution

$$E_{wc} = E_{lim} \frac{I_T}{I_{lim}} \left[ 1 - \left( \frac{I_T}{I_{lim}} \right)^2 \right]^{-1/2} \leq E_{lim}. \quad (3.20)$$

The square root in the denominator represents the fact that with increasing current density and field the mobility decreases due to its field dependence, which has to be compensated for by an overproportional increase in  $E_{wc}$  in order to keep carrying the current density. Note, that the assumption of ohmic current flow with the resistivity given by the collector(-epi) doping at  $x = w_{Ci}$  restricts the application of the above equation to current densities below  $I_{lim}$ .

The next step is to calculate the remaining two variables, i.e.  $E_{jc}$  and  $w_{BC,c}$ . The relation between field and applied voltage is given by integrating Poisson's equation (3.3) twice. The first integration over the space charge region gives the spatially dependent field (3.13). Integrating again yields with  $E(0) = -E_{jc}$

$$\int_0^{w_{BC,c}} E dx = -E_{jc} w_{BC,c} + \frac{qN_{Ci}}{2\varepsilon} w_{BC,c}^2 \left( 1 - \frac{I_T}{I_{lim}} \right), \quad (3.21)$$

while integration over the ohmic region leads to

$$\int_{w_{BC,c}}^{w_{Ci}} E dx = -E_{wc} (w_{Ci} - w_{BC,c}). \quad (3.22)$$

The total effective collector voltage is then the sum of the integrals over the field in both regions:

$$V_{ci} = E_{jc} w_{BC,c} - \frac{qN_{Ci}}{2\varepsilon} w_{BC,c}^2 \left( 1 - \frac{I_T}{I_{lim}} \right) + E_{wc} (w_{Ci} - w_{BC,c}). \quad (3.23)$$

The second boundary condition is given by the intercept of the field curves for space charge and ohmic region at  $w_{BC,c}$ , as shown in the idealized curve in Fig. 3.7:

$$E(w_{BC,c}) = -E_{jc} + \frac{qN_{Ci}}{\varepsilon} \left( 1 - \frac{I_T}{I_{lim}} \right) w_{BC,c} = -E_{wc}. \quad (3.24)$$

Solving above equation for

$$E_{jc} = E_{wc} + \frac{qN_{Ci}}{\varepsilon} \left(1 - \frac{I_T}{I_{lim}}\right) w_{BC,c} \quad (3.25)$$

and inserting the expression into (3.23) yields the lengthy expression

$$V_{ci} = E_{wc} w_{BC,c} + \frac{qN_{Ci}}{\varepsilon} \left(1 - \frac{I_T}{I_{lim}}\right) w_{BC,c}^2 - \frac{qN_{Ci}}{2\varepsilon} w_{BC,c}^2 \left(1 - \frac{I_T}{I_{lim}}\right) + E_{wc} (w_{Ci} - w_{BC,c}).$$

Cancelling and arranging terms properly gives

$$V_{ci} = \frac{qN_{Ci}}{2\varepsilon} \left(1 - \frac{I_T}{I_{lim}}\right) w_{BC,c}^2 + E_{wc} w_{Ci}. \quad (3.26)$$

This leads to a fairly simple equation for the bias dependent SCR width:

$$w_{BC,c} = w_{Ci} \sqrt{\frac{V_{ci} - E_{wc} w_{Ci}}{V_{PT}(1 - I_T/I_{lim})}}. \quad (3.27)$$

The form of this expression is exactly the same as the one obtained in [10], except that the field dependence of the mobility in the non-depleted ohmic region (cf. (3.20)) was neglected in [10]. Nevertheless, the corresponding theory is valid under the same assumptions and still corresponds to the idealized field distribution shown in Fig. 3.7. Inserting  $w_{BC,c}$  into (3.25) gives the desired bias dependent electric field at the BC junction,

$$E_{jc} = E_{wc} + \sqrt{\frac{2qN_{Ci}}{\varepsilon}} \sqrt{\left(1 - \frac{I_T}{I_{lim}}\right) (V_{ci} - E_{wc} w_{Ci})} > 0 \quad (3.28)$$

with  $E_{wc}$  from (3.20).

Next, the bias limits, within which the above equation can be used, and their physical meaning are discussed. Obviously, there is no limit towards low current densities, and for  $I_T = 0$  the field assumes its low-current value

$$E_{jc0} = \sqrt{\frac{2qN_{Ci}}{\varepsilon}} V_{DCi} \sqrt{\frac{V_{ci}}{V_{DCi}}} = E_{jc00} \sqrt{\frac{V_{ci}}{V_{DCi}}}. \quad (3.29)$$

where  $E_{jc00} = E_{jc}(V_{B'C'} = 0, I_T = 0)$ . The first limit exists, if at  $I_T = 0$  the applied (reverse) voltage is large enough to fully deplete the epi region. However, the according condition also occurs at non-zero current density in a similar way and, hence, the criterion can be generalized using (3.27),

$$w_{BC, c}|_{PT} = w_{Ci} \sqrt{\frac{V_{ci, PTJ} - E_{wc} w_{Ci}}{V_{PT}(1 - I_T/I_{lim})}} = w_{Ci}, \quad (3.30)$$

with the punch-through voltage  $V_{PT}$  at  $I_T = 0$  from (3.9). To guarantee above condition, the resulting maximum voltage, up to which (3.28) can be applied is current dependent and reads

$$V_{ci, PTJ}(V_{B'C}, J_T) \leq V_{PT} \left(1 - \frac{I_T}{I_{lim}}\right) + E_{wc} w_{Ci}. \quad (3.31)$$

The physical interpretation of above relation is as follows. The first term represents the punch-through voltage that would exist if the voltage drop across the remaining epi region, which is represented by the second term and reduces the reverse voltage across the SCR, would be negligible. The current dependent factor in the first term represents the compensation of the ionized atoms in the SCR by mobile carriers. For  $V_{ci} > V_{ci, PTJ}(I_T)$  equation (3.28) has to be replaced by the corresponding punch-through equation (3.14) for  $E_{jc}$ . Note, that there is again a small region at the end of the SCR in which the electric field may not be large enough to support velocity saturation, and the respective assumption is violated. This inconsistency could be eliminated by using the condition  $E_{wc} = E_{lim}$  for a fully depleted collector. The other limitation for the validity of (3.28) occurs below a certain value of  $V_{ci, 0}$  when the BC SCR collapses. The corresponding mathematical condition is given by

$$w_{BC, c} = w_{Ci} \sqrt{\frac{V_{ci, 0} + E_{wc} w_{Ci}}{V_{PT}(1 - I_T/I_{lim})}} \geq 0 \quad (3.32)$$

and results in

$$V_{ci} \geq E_{wc}(J_T) w_{Ci}. \quad (3.33)$$

Physically, the "=" sign corresponds to a spatially independent electric field in the entire epi-collector. According to (3.28), the square root becomes zero, and  $E_{jc} = E_{wc} = V_{ci}/w_{Ci}$ . Since  $E_{wc}$  can assume at maximum the value  $E_{lim}$  (otherwise the region would be a SCR, see also (3.20)), the field occurring at the "=" condition in (3.33) is too small to support velocity saturation. This is consistent with the original assumption for defining  $E_{wc}$  as the field in an ohmic region. In other words, the collector becomes

ohmic. Therefore, if for a given  $V_{ci}$  (i.e.  $V_{B'C'}$ ) the current causes the above condition to fail equation (3.28) has to be replaced by an equation that describes an *ohmic* collector region at the BC junction.

If at low current densities the collector voltage is not large enough to cause punch-through, then the expressions derived for the low-voltage case have to be applied. Once at larger currents the punch-through occurs, the above expressions for the high-voltage case apply. Since the transition between the expressions is not always continuously differentiable the equations are not suitable for a compact model in their present form. However, they can serve as a basis for a smooth and fairly simple analytical description of the electric field as a function of bias [12].

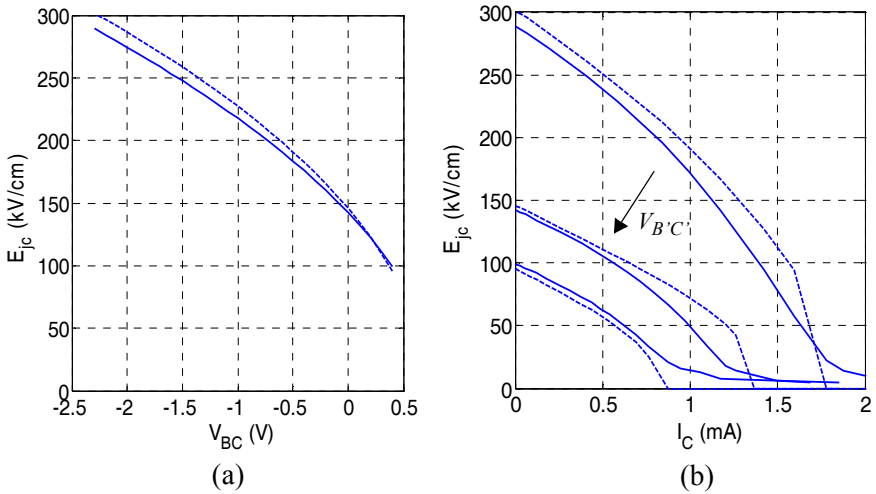


Fig. 3.8: Bias dependence of the electric field at the BC junction of the BJT. Comparison between device simulation (solid lines) and theory (dashed lines): (a) voltage dependence (3.29) at low injection and (b) current dependence (3.28) with (3.20) for different voltages  $V_{B'C'}/V = 0.4, 0, -2.3$ .



### 3.3 Static operation and characteristics

This chapter contains a derivation of basic relations between currents and applied voltages in a 1D npn bipolar transistor for the DC case, which is defined by a negligible time derivative in the continuity equations (i.e.  $\partial/\partial t = 0$  in (2.8b)). First, section 3.3.1 presents conventional theories for the DC transfer current, which is the most important component of the collector terminal current. In section 3.3.2, the base current components are discussed that result from back injection across the junctions. Then, the avalanche current caused by base-collector breakdown is considered in section 3.3.4, followed in section 3.3.3 by a derivation of the current components resulting from recombination in the SCR.

The discussion and all results can be applied in a similar way to pnp transistors. Each section discusses the assumptions required for the derivation of the analytical relations. Their accuracy is then assessed through comparisons with device simulation.

#### 3.3.1 Transfer current

The transfer current is usually understood as the electron current reaching the internal collector contact. For a 1D structure this corresponds to the relation

$$I_T = -J_{nx}(x_C)A_E \quad (3.34)$$

where  $J_{nx}(x_C)$  is the (electrical) electron current density at the internal collector contact, and  $A_E$  is the emitter area which equals the unit area in the 1D case.

In the classical literature (e.g. [1,2,4]) just the neutral base region is considered for deriving the transfer current relation. Including recombination in the continuity equation, a second-order differential equation for the static minority carrier density in the neutral base is then obtained. The solution is rather lengthy, and the result is unnecessarily complicated for applications in modern transistors as discussed below.

It was shown in section 3.1, that in existing bipolar transistors the influence of recombination on the electron current density is negligible within the base and collector region. In the emitter region, this assumption is also

valid for sufficiently high current gain, which is the case in most Si and SiGe transistors in the practically most relevant operating region, i.e. at not too low or high injection. Furthermore, defining the transfer current more precisely as the fundamental quasi-static current component consisting of electrons injected by the emitter, passing through the base and collector with negligible recombination, and neglecting avalanche generation for the time being, (2.8b) in such regions simply reads

$$\frac{dJ_{nx}}{dx} = 0 \quad (3.35)$$

which corresponds to a *spatially independent*  $J_{nx}$  value:

$$J_{nx} = \frac{-I_T}{A_E}. \quad (3.36)$$

As a consequence, only the transport equation (2.9b) has to be solved, which is a *first-order* differential equation for the minority carrier density. This not only results in a much simpler solution procedure along the lines of classical theory (labelled here as "conventional" approach) but also enables a solution path towards a much more general formulation of the transfer current (cf. ch. 4). The conventional approach is considered below assuming, for the time being, that effects such as thermionic emission and tunneling across the junctions can be neglected. This is a valid assumption for Si based transistors, but is not applicable to certain types of III-V HBTs. However, the goal at this point is to provide the basic set of relations that can be employed in a simple circuit design-oriented compact model. Additional effects can be accounted for later.

Under the assumptions mentioned before, only the transport equation needs to be solved for the neutral base region with width  $w_B$  and doping concentration  $N_B$ . However, solving the differential equation (2.9a) *analytically* in a *closed-form suitable for compact models* still requires some additional simplifications that depend on the operating range and will be discussed in the subsections below. A relation for the transfer current will be derived first for the case of low injection and then for high injection at a given arbitrary bias point in the respective operating region. Afterwards, the limitations of both solutions are discussed along with the case of me-

dium injection, and consequences for the practical application of the solutions will be pointed out.

### 3.3.1.1 Low injection

This operating region is characterized by the conditions

$$n \ll N_B \quad \text{and} \quad p \cong N_B. \quad (3.37)$$

Under low injection  $J_{px}$  in (2.9a) is negligible compared to the respective diffusion and drift component. Hence, applying condition (3.37), the effective electric field in (2.9a) can be expressed as

$$E_{px} = \frac{V_T}{N_B} \frac{dN_B}{dx}, \quad (3.38)$$

from which, according to (2.12a), the electrostatic field follows:

$$E_x = E_{px} - \frac{dV_p}{dx} = \frac{V_T}{N_B} \frac{dN_B}{dx} - \frac{dV_p}{dx}. \quad (3.39)$$

Inserting this into (2.12b) gives the effective field for electrons,

$$E_{nx} = \frac{V_T}{N_B} \frac{dN_B}{dx} - \frac{d(V_p + V_n)}{dx}. \quad (3.40)$$

According to (2.7) the sum of the band potentials contains all built-in contributions to a change in band edges and is, for a BJT, directly related to the intrinsic carrier density,

$$V_n + V_p = 2V_T \ln \frac{n_i}{n_{ir}}. \quad (3.41)$$

Hence, one can also write

$$E_{nx} = \frac{V_T}{N_B} \frac{dN_B}{dx} - 2 \frac{V_T}{n_i} \frac{dn_i}{dx}. \quad (3.42)$$

For many practical cases, an exponential doping profile in the neutral base region and a linear change in band edges is a good first-order approximation. This leads to a spatially constant field  $\bar{E}_{nx}$ . Therefore, an average effective field is a reasonable representation for the realistic case.

To bring (2.9b) into a tractable form for a closed-form analytical solution, the further assumption of a spatially independent average electron mobility,  $\bar{\mu}_{nB}$ , is being made. The starting equation then reads

$$J_{nx} = q\bar{\mu}_{nB}n\bar{E}_{nx} + q\bar{\mu}_{nB}V_T\frac{dn}{dx}, \quad (3.43)$$

which yields after inserting (3.36) and rearrangement of terms a first-order differential equation for the minorities in the neutral base:

$$\frac{dn}{dx} + \frac{\bar{E}_{nx}}{V_T}n = \frac{-I_T}{qA_E\bar{\mu}_{nB}V_T}. \quad (3.44)$$

The solution for the spatially dependent electron density is

$$n(x) = \frac{I_T w_B}{qA_E\bar{\mu}_{nB}V_T\zeta} \left[ 1 - \exp\left(-\zeta\left(1 - \frac{x-x_e}{w_B}\right)\right) \right] + n_c \exp\left(-\zeta\left(1 - \frac{x-x_e}{w_B}\right)\right) \quad (3.45)$$

with the so-called drift factor

$$\zeta = -\bar{E}_{nx}w_B/V_T, \quad (3.46)$$

$x_e$  as the emitter-sided edge of the neutral base (cf. (2.24)) and

$$n_c = n(x_c) \quad (3.47)$$

as electron density at the collector-sided edge  $x_c$  of the neutral base.

From (3.45), the electron density at the  $x_e$  is given by

$$n_e = n(x_e) = \frac{I_T w_B}{qA_E\bar{\mu}_{nB}V_T\zeta} [1 - \exp(-\zeta)] + n_c \exp(-\zeta). \quad (3.48)$$

Defining the drift function

$$f_\zeta = \exp(\zeta) \quad (3.49)$$

and solving for the transfer current yields

$$I_T = A_E \frac{q\bar{\mu}_{nB}V_T}{w_B} \cdot \frac{\zeta \cdot f_\zeta}{f_\zeta - 1} \left( n_e - \frac{n_c}{f_\zeta} \right). \quad (3.50)$$

The relation between  $I_T$  and the controlling terminal voltages can be obtained from the bias dependence of  $n_e$  and  $n_c$ . Using the pn product at  $x_e$ ,

$$p(x_e) n(x_e) = n_i^2(x_e) \exp\left(\frac{\Phi_p(x_e) - \Phi_n(x_e)}{V_T}\right), \quad (3.51)$$

and inserting  $p$  from this equation into the neutrality condition at  $x_e$ ,

$$p(x_e) - n(x_e) - N_B(x_e) = 0, \quad (3.52)$$

gives a quadratic equation for the electron density the solution of which is

$$n_e = \frac{N_B(x_e)}{2} \left[ \sqrt{1 + 4 \frac{n_i^2(x_e)}{N_B^2(x_e)} \exp\left(\frac{V_{B'E'}}{V_T}\right)} - 1 \right]. \quad (3.53)$$

Here, the difference of the quasi-Fermi potentials was replaced by the internal BE voltage,

$$V_{B'E'} = \phi_p(x_e) - \phi_n(x_e), \quad (3.54)$$

which was shown in section 3.1 to be a valid assumption. Under the low injection conditions assumed here (3.53) simplifies to

$$n_e = \frac{n_i^2(x_e)}{N_B(x_e)} \exp\left(\frac{V_{B'E'}}{V_T}\right). \quad (3.55)$$

Figure 3.9 shows the voltage dependence of  $n_e$  calculated from (3.53) and (3.55) with  $N_B$  taken at  $x_e(0.7V)$ , and normalized to its value at  $V_{B'E'} = 0.7V$ . Since it is difficult to obtain  $x_e$  as *smooth* function of bias (e.g. using the Regional Approach), instead of the voltage dependent  $n_e$  the collector current  $I_C$  has been inserted directly for comparison. At low injection,  $I_C$  follows quite well the carrier density. However, at high current densities (3.53) starts to deviate significantly, which is caused by the fact that (3.53) only includes base conductivity modulation through the second term under the square root in (3.53). In contrast the strong change in the slope of  $I_C$  results from high-current effects in the *collector*, which are indicated by the sudden increase of the electron density  $n_c$  that is also shown in Fig. 3.9. Note that the (ideal) low-injection approximation (3.55) shows very large deviations at high injection.

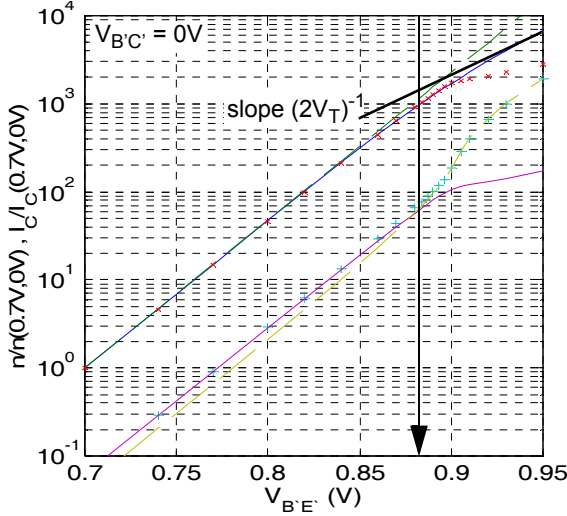


Fig. 3.9: Normalized electron density at  $x_e(V_{B'E'} = 0.7V) = 0.028\mu\text{m}$  and  $x_c(V_{B'E'} = 0.7V, V_{B'C'} = 0V) = 0.055\mu\text{m}$  vs. BE voltage. Comparison for  $n_e$  (upper curves) between eq. (3.53) (solid line), eq. (3.55) (dashed line) and the normalized collector current (crosses). Comparison for  $n_c$  (lower curves) between eq. (3.57) (dashed line), eq. (3.58) (solid line) and normalized  $n_c(x_c)$  from device simulation (crosses). The arrow indicates the location of peak  $f_T$ .

Analogous relations can be derived for  $n_c$  if  $x_e$  and  $n_e$ , respectively, are replaced by  $x_c$  and  $n_c$ , respectively, and if  $V_{B'E'}$  is replaced by

$$V_{BCi} = \phi_p(x_c) - \phi_n(x_c). \quad (3.56)$$

This voltage should not be confused with the terminal voltage  $V_{B'C'}$  or the voltage drop across the BC junction as is evident from Fig. 3.10a. At low current densities  $V_{BCi}$  just follows  $\phi_p(x_c) = V_{B'E'}$  and is reduced by a bias independent  $\phi_n(x_c) > 0$ . Towards high current densities, the electron density associated with the current through the BC SCR leads to a drop of the electric field and also of  $\phi_n(x_c)$ , resulting in a more than proportional increase of  $V_{BCi}$  with  $V_{B'E'}$ .

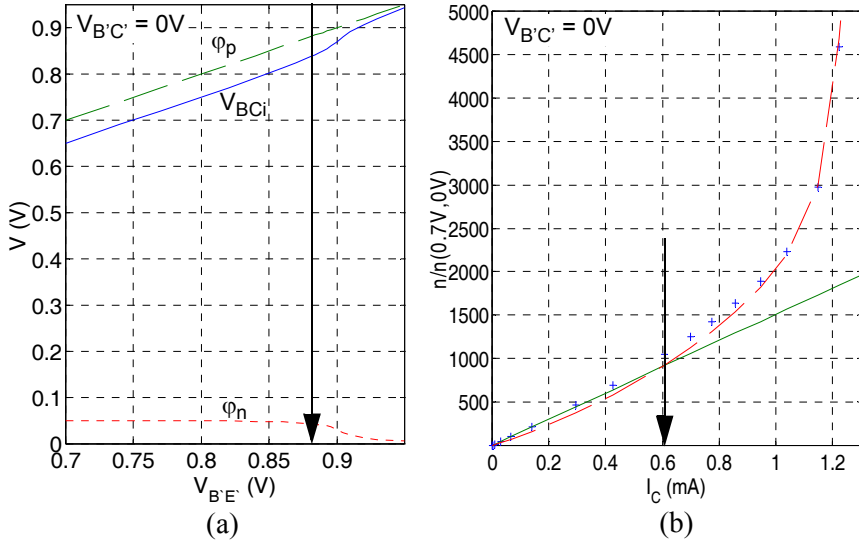


Fig. 3.10: (a) Potentials vs. BE voltage:  $V_{BCi}$  from (3.56), quasi-Fermi potentials  $\phi_p(x_{jc})$  and  $\phi_n(x_{jc})$  at  $V_{B'C'} = 0V$ . (b) Normalized electron density  $n_c$ , taken at the end of the neutral base ( $x_c = 0.055\mu m$  at  $V_{B'E} = 0.7V$ ), as function of the collector current  $I_c$ . Comparison between device simulation (symbols), eq. (3.57) (dashed line) and eq. (3.58) (solid line). The arrow indicates the location of peak  $f_T$ .

From an application point of view,

$$n_c = \frac{N_B(x_c)}{2} \left[ \sqrt{1 + 4 \frac{n_i^2(x_c)}{N_B^2(x_c)} \exp\left(\frac{V_{BCi}}{V_T}\right)} - 1 \right]. \quad (3.57)$$

can only be calculated if  $V_{BCi}$  is related to a terminal voltage. Modeling the bias dependence of  $V_{BCi}$  directly is very difficult though since very high accuracy is required due to the exponential dependence of  $n_c$  on  $V_{BCi}$ . One solution would be a separate relation for the current through the collector region, which uses  $V_{BCi}$  to force this current to be equal to  $I_T$  from, e.g., (3.50) [13, 14]. As shown in Fig. 3.10b, (3.57) is a quite accurate description if an accurate analytical expression for  $V_{BCi}$  is available.

In classical theory  $n_c = 0$  is assumed, which corresponds to the non-physical assumption of infinite electron drift velocity in the BC SCR. However, under the assumption of negative or, at maximum, weak forward

$V_{B'C'}$  and low injection (i.e. negligible impact of  $J_{nx}$  on the BC space-charge),  $n_c$  is given simply by drift of carriers through the BC SCR,

$$n_c = \frac{J_{nx}}{qv_c} \cong \frac{I_{Tf}}{A_E q v_c} . \quad (3.58)$$

Here,  $v_c$  is the field dependent electron drift velocity, and the reverse transfer current is negligible compared to the forward component  $I_{Tf}$ . As is evident from Fig. 3.10b,  $n_c$  (normalized) is quite well described by (3.58) at low injection using a constant velocity. However, once at higher current densities the electric field in the BC SCR drops, the velocity needs to be modelled as function of  $I_C$ .

For practical purposes and further discussions, various factors can be defined in (3.50). Common to both carrier densities is the “unity” diffusion current density

$$J_{Tdl} = \frac{q \bar{\mu}_{nB} V_T}{w_B} , \quad (3.59)$$

while the drift field acts differently upon  $n_e$  and  $n_c$ , which is expressed by the factors

$$a_{Bfl} = \frac{f_\zeta - 1}{f_\zeta \zeta} , \quad a_{Brl} = \frac{f_\zeta - 1}{\zeta} . \quad (3.60)$$

Classical theory separates the transfer current into a forward and reverse component,

$$I_T = I_{Tf} - I_{Tr} . \quad (3.61)$$

With the previously defined factors, the two components read

$$I_{Tf} = A_E \frac{J_{Tdl}}{a_{Bfl}} n_e , \quad (3.62)$$

$$I_{Tr} = A_E \frac{J_{Tdl}}{a_{Brl}} n_c . \quad (3.63)$$

Inserting  $n_e$  from (3.55) and the corresponding relation for  $n_c$  then yields  $I_T$  as function of controlling voltages. For a negligible effective field, the usual textbook solution of a diffusion transistor is obtained. Letting  $\zeta \rightarrow 0$



gives for the electron density in the neutral base the well-known linear spatial dependence,

$$n(x) = \frac{I_T w_B}{q A_E \mu_{nB} V_T} \left( 1 - \frac{x - x_e}{w_B} \right) + n_c \quad (3.64)$$

and, with  $a_{Bfl} = 1$  and  $a_{Brl} = 1$ , for the transfer current

$$I_T = I_{Tf} - I_{Tr} = A_E J_{Tdl} n_e - A_E J_{Tdl} n_c. \quad (3.65)$$

### 3.3.1.2 Very high injection

This operating region is characterized by the condition

$$n \cong p \gg N_B. \quad (3.66)$$

Since the electron density in the base now influences the hole density, the hole transport equation needs to be considered, too (Webster effect [15]). Applying above condition and assuming again  $J_{px}$  in (2.9a) to be negligible, the effective electric field now reads

$$E_{px} = \frac{V_T dp}{p dx} \cong \frac{V_T dn}{n dx}. \quad (3.67)$$

Inserting again the electrostatic field from (2.12a) into (2.12b) gives the effective field for electrons,

$$E_{nx} = \frac{V_T dn}{n dx} - \frac{d(V_p + V_n)}{dx}. \quad (3.68)$$

While the first term is now bias dependent, the second term remains bias independent and contains the built-in changes of the band edges. Defining the band-gap related built-in field as

$$E_{bg} = -\frac{1}{2} \frac{d(V_p + V_n)}{dx}, \quad (3.69)$$

inserting above field relations into (2.9b) gives

$$J_{nx} = q \mu_n n \frac{V_T dn}{n dx} + q \mu_n V_T \frac{dn}{dx} + 2 q \mu_n n E_{bg}. \quad (3.70)$$

Note that if  $J_{px}$  cannot be neglected, an ambipolar transport equation results, which is of little use for BJT compact modeling.

Sorting terms and assuming an average value  $\bar{E}_{bg}$  in the neutral base leads to<sup>1</sup>

$$J_{nx} = 2 \left( q \bar{\mu}_{nBh} V_T \frac{dn}{dx} + q \bar{\mu}_{nBh} n \bar{E}_{bg} \right). \quad (3.71)$$

Here also a spatial average  $\bar{\mu}_{nBh}$  for the mobility is assumed, which is different from the low-current value due to the very different electric field and carrier densities. Rearranging terms yields the same form as (3.44):

$$\frac{dn}{dx} + \frac{\bar{E}_{bg}}{V_T} n = \frac{I_T}{2qA_E \bar{\mu}_{nBh} V_T}. \quad (3.72)$$

Therefore, the solution for  $n(x)$  has the same form as (3.45), *except for a factor 2* in front of the mobility,

$$n(x) = \frac{I_T w_B}{2qA_E \bar{\mu}_{nBh} V_T \zeta_h} \left[ 1 - \exp \left( -\zeta_h \left( 1 - \frac{x-x_e}{w_B} \right) \right) \right] + n_c \exp \left( -\zeta_h \left( 1 - \frac{x-x_e}{w_B} \right) \right), \quad (3.73)$$

and the high-injection drift factor

$$\zeta_h \cong -\bar{E}_{bg} w_{Bm} / V_T, \quad (3.74)$$

in which the SCR widths in the base were neglected. Following the same procedure as for low injection, the resulting transfer current now reads

$$I_T = I_{Tf} - I_{Tr} = A_E \frac{J_{TDh}}{a_{Bfh}} n_e - A_E \frac{J_{TDh}}{a_{Brh}} n_c \quad (3.75)$$

with the high injection related unity diffusion current density

$$J_{TDh} = 2 \frac{q \bar{\mu}_{nBh} V_T}{w_B} \quad (3.76)$$

and the drift field related factors

---

1. Generally, the current under base conductivity modulation is determined by ambipolar transport [4].

$$a_{Bfh} = \frac{f_{\zeta_h} - 1}{f_{\zeta_h} \zeta_h}, \quad a_{Brh} = \frac{f_{\zeta_h} - 1}{\zeta_h}. \quad (3.77)$$

The remaining built-in drift field still acts differently upon  $n_e$  and  $n_c$ . In contrast to low injection though different boundary conditions for the carrier densities apply. For  $n_e$ , the voltage dependent exponential term in (3.53) now dominates, leading to the asymptotic dependence

$$n_e = n_i(x_e) \exp\left(\frac{V_{BE}}{2V_T}\right). \quad (3.78)$$

As can be observed in Fig. 3.9, the corresponding slope of  $(2V_T)^{-1}$  is significantly too large so that the classical theory, with its limitation to the base region, cannot be applied in practice.

For  $n_c$ , the assumption of a drift current at the BC junction is not justified anymore if the collector doping is lower than the base doping, which is the case for all practical transistors. Therefore, eq. (3.57) has to be used with  $V_{BCi}$  defined by (3.56) and  $x_c \approx x_{jC}$ . As shown before in Figs. 3.9 and 3.10,  $n_c$  can be described accurately if the exact values from device simulation are inserted for  $V_{BCi}$ . For a compact model though  $V_{BCi}$  needs to be related analytically to the terminal voltage. A possible solution was given in [13, 14], which is limited though to thick collector epi regions and "low" voltages  $V_{CE}$ . Semi-empirical extensions to realistic cases can be found in [16] and related literature.

In classical theory, and even in newer literature [17, 5], the above analysis is performed with  $n_c = 0$  and  $E_{bg} = 0$ . In case of a negligible drift field, the solution of a diffusion transistor is recovered from (3.73) for the electron density in the neutral base:

$$n(x) = \frac{I_T w_B}{2q A_E \bar{\mu}_{nBh} V_T} \left(1 - \frac{x - x_e}{w_B}\right) + n_c. \quad (3.79)$$

As compared to low injection the spatial slope is now reduced by a factor 2 which reflects the weaker influence of the control voltage on  $n_e$ . With  $a_{Bfh} = 1$  and  $a_{Brh} = 1$ , the transfer current reads for a diffusion transistor

$$I_T = I_{Tf} - I_{Tr} = A_E J_{TDh} n_e - A_E J_{TDh} n_c. \quad (3.80)$$

### 3.3.1.3 Medium injection

For medium injection, neither the assumption (3.37) nor (3.66) is valid, leading to an ambipolar transport equation that does not have a closed-form analytical solution. An often pursued idea is to connect the solutions obtained so far for the asymptotic cases of (very) low and (very) high injection through a smoothing function. Summarizing the results for the transfer current components yields

$$i_{Tf} = A_E n_e \begin{cases} J_{TDl}/a_{Bfl} & , \quad n \ll N_B \\ J_{TDh}/a_{Bfh} & , \quad n \gg N_B \end{cases} \quad (3.81)$$

and

$$i_{Tr} = A_E n_c \begin{cases} J_{TDl}/a_{Br l} & , \quad n \ll N_B \\ J_{TDh}/a_{Br h} & , \quad n \gg N_B \end{cases} \quad (3.82)$$

Even, if the general relations (3.53) and (3.57) for the voltage dependence were used and a pure diffusion transistor were considered, two different solutions are obtained because of the remaining factor 2 difference between  $J_{TDl}$  and  $J_{TDh}$ . Therefore, there is no simple way of combining the conventional solution into a single compact equation for the transfer current. Using proper smoothing functions though, a single-piece formulation can be obtained (e.g. [18]). However, there is still the remaining problem of how to describe the bias dependence of  $n_c$  at medium to high current densities, which requires analytical solutions for the carrier and current density in the (more lightly doped) collector region under all possible bias conditions.

The derivation of the above equations was performed for a single bias point, albeit different for the low and high injection case. This already highlights another issue: the neutral base boundaries  $x_e$  and  $x_c$  are bias dependent (cf. sec. 2.4.1)! For instance, the bias dependence of  $x_c$  represents the forward Early-effect. Hence, a practically useful compact model requires the formulation of  $x_e$  and  $x_c$  as a function of bias (and temperature, too). Such a description is usually being added to an existing transfer current relation as an afterthought, but is not easy to set up accurately, since the boundary conditions (3.53) and (3.57) also contain  $x_e$  and  $x_c$ .

A comparison of (3.61)-(3.63) and (3.75) with 1D device simulation in Fig. 3.11 reveals the difficulties of conventional theory to accurately describe modern transistors. For this comparison,  $V_{B'C'} = 0$  is sufficient to demonstrate the main issues just for forward operation. With (3.55) and (3.58) the transfer current components at low injection can be written as

$$I_{Tfl} = A_E \frac{J_{TDl} n_i^2(x_e)}{a_{Bfl} N_B(x_e)} \exp\left(\frac{V_{B'E'}}{V_T}\right) = I_{Sfl} \exp\left(\frac{V_{B'E'}}{V_T}\right), \quad (3.83)$$

$$I_{Trl} = \frac{J_{TDl}}{a_{Brl} q v_{ns}} I_{Tfl} = C_{Trl} I_{Tfl}. \quad (3.84)$$

While  $I_{Sfl}$  has been determined at  $V_{B'E'} = 0.7V$ ,  $C_{Trl}$  cannot be determined unless  $I_{Trl}$  significantly impacts the total transfer current. This is only the case at either high injection or very low CE voltages. At both conditions, relation (3.58) becomes invalid, so that (3.83) is sufficient to describe the low injection case, which is then equivalent to the condition  $I_{Tf} \gg I_{Tr}$ . Figure 3.11 shows that  $I_{Tf}$  is a good approximation at low injection. Note, that  $I_{Sfl}$  is a constant only if  $x_e$  and  $x_c$  and, as a result,  $w_B$  are bias independent; i.e. if the Early-effect in both forward and reverse operation is neglected. Hence, the error is larger for voltages other than  $V_{B'C'} = 0V$ .

At high injection, one needs to switch to different relations with different parameters that read

$$I_{Tfh} = A_E \frac{J_{TDh}}{a_{Bfh}} n_i(x_e) \exp\left(\frac{V_{B'E'}}{2V_T}\right) = I_{Sfh} \exp\left(\frac{V_{B'E'}}{2V_T}\right), \quad (3.85)$$

$$I_{Trh} = A_E \frac{J_{TDh}}{a_{Brh}} n_i(x_c) \exp\left(\frac{V_{BCi}}{2V_T}\right) = I_{Srh} \exp\left(\frac{V_{BCi}}{2V_T}\right). \quad (3.86)$$

Here, (3.57) has been simplified similarly to the forward term, assuming the voltage dependent term under the square root dominates. For the comparison with device simulation in Fig. 3.11, the Parameters  $I_{Sfh}$  and  $I_{Srh}$  were fitted, and (3.56) was used for  $V_{BCi}$  with quasi-Fermi-potentials taken directly from device simulation. As can be seen,  $I_{Tfh}$  agrees with  $I_C$  only for a very small region due to the high-current effects in the collector which are not included. From the behavior of  $I_{Trh}$  in Fig. 3.11 it is also obvious that the current  $I_{Th} = I_{Tfh} - I_{Trh}$  becomes even negative at high injection.

tion. In other words, the classical concept of superimposing a forward and reverse current component, calculated from transport in the base region, does not work for actual transistors with collector doping lower than base doping.

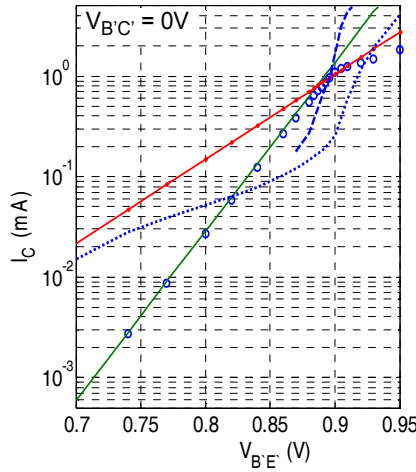


Fig. 3.11: Forward transfer current vs. BE voltage at  $V_{B'C'} = 0$ : comparison between  $I_C$  from 1D device simulation (crosses), eq. (3.83) for  $I_{Tn}$  (solid line), eq. (3.85) for  $I_{Tn}$  (solid line with dots) and eq. (3.86) for  $I_{Trh}$  (dashed line). In addition, the relative error  $|1 - I_{Tn}/I_C|$  is shown as dotted line.

### 3.3.2 Back injection related current components

In the forward-bias operating range relevant for most circuit applications recombination in the SCRs and neutral base does not play a significant role. As a consequence, the corresponding  $I_B(V_{B'E'})$  characteristic in modern transistors is determined mainly by the injection of holes into the emitter and, hence, recombination in the neutral mono-silicon and poly-silicon emitter region as well as at the poly-silicon to mono-silicon interface. Figure 3.12 shows a sketch of the hole distribution in the entire emitter region and the respective dimensions used for the calculations below. In all modern technologies the mono-silicon width  $w_E$  is small compared to the poly-silicon width  $w_P$ . Between these two layers often a thin interfacial oxide exists that can cause a recombination current density

$$J_{pmI} = q\Delta p_m(0)v_{pmI} \quad (3.87)$$

with  $v_{pmI}$  as the interface recombination velocity. The sketch in Fig. 3.12 shows only a single poly-silicon region while in reality more than one section exists depending on the grain size. Traps at each grain boundary can contribute to recombination current densities that can be described similarly to (3.87). Within the grains volume recombination occurs that can be treated the same way as in bulk silicon but with a smaller diffusion length.

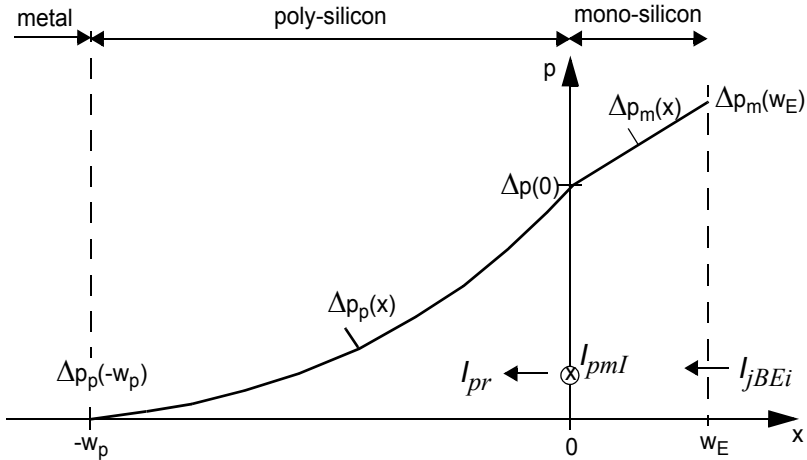


Fig. 3.12: Sketch of the hole density distribution in the emitter region and corresponding hole currents.  $I_{pmI}$  and  $I_{pr}$  are the interface and poly recombination currents.  $w_E$  and  $w_p$  are the widths of the mono- and poly-silicon region. Tunnelling through the interface is neglected. The carrier densities indicated will be needed later in the charge calculation.

There have been many investigations in the literature about the carrier transport at the poly-to-mono-silicon interface and in poly-silicon. However, most results are not suitable for compact models. This is partially due to the fact that the material properties of poly-silicon, such as grain size and carrier lifetime, depend to a large extent on the processing conditions making accurate theoretical predictions very difficult. The purpose of this section is to provide practically useful relations with a reasonable physics-based background for calculating the hole current in the neutral emitter region. First, important properties of the emitter region in modern bipolar transistors are reviewed that allows certain simplifications of the basic

equations. Then a suitable analytical solution is given for the base current that includes the most relevant physical parameters of the emitter region.

For a general analysis, recombination can not be neglected. However, the recombination term can be simplified to

$$R = \frac{p - p_0}{\tau_{pE}} \quad (3.88)$$

with  $\tau_{pE}$  as an average recombination time constant and  $p_0$  as the hole density at thermal equilibrium. Furthermore, the time derivative can be neglected for quasi-static calculations, i.e.  $\partial/\partial t = 0$ . With  $N_E$  as the doping concentration in the neutral emitter and making the justified assumption that low injection prevails in the emitter under all practical bias conditions, the hole density in thermal equilibrium is given by

$$p_0 = n_{iE}^2 / N_E. \quad (3.89)$$

The effective electric field in the transport equation can be calculated from the equilibrium hole density,

$$p_0 = n_{i0} \exp\left(\frac{V_p - \psi}{V_T}\right), \quad (3.90)$$

by taking the derivative which gives

$$\frac{d(V_p - \psi)}{dx} = \frac{dV_p}{dx} + E_x = E_{px} = \frac{V_T dp_0}{p_0 dx}. \quad (3.91)$$

Inserting (3.89) gives for the effective field the general expression

$$E_{px} = \frac{V_T dp_0}{p_0 dx} = \frac{V_T dn_{iE}^2}{n_{iE}^2 dx} - \frac{V_T dN_E}{N_E dx}. \quad (3.92)$$

Here, the  $N_E$  term includes a possible spatial dependence of the emitter doping while the  $n_{iE}^2$  term contains all bandgap related effects. Since practical Si-based transistors do not contain any emitter material composition,  $n_{iE}^2$  is determined by high-doping effects only. Using the approximation given in [19] leads for typical emitter doping concentrations  $N_E \gg D_{ni} = 2 \cdot 10^{17} \text{ cm}^{-3}$  to



$$n_{iE}^2 \approx n_{i0}^2 \left( \frac{N_E}{D_{ni}} \right)^{2V_{ni}/V_T}. \quad (3.93)$$

Inserting above relation in (3.92) yields

$$E_{px} \cong 2V_{ni} \frac{1}{N_E} \frac{dN_E}{dx} - \frac{V_T}{N_E} \frac{dN_E}{dx} = (2V_{ni} - V_T) \frac{1}{N_E} \frac{dN_E}{dx}, \quad (3.94)$$

With  $V_{ni} \approx 9\text{mV}$ , this gives about  $8\text{kV/cm}$  for a doping gradient of  $10^{20}\text{cm}^{-3}/10\text{nm}$ , and is still a moderate value for a quite large gradient. Therefore, and since the high emitter doping prevents any larger ohmic voltage drop, the effective electric field is neglected and so is the field dependence of the hole mobility. Therefore, a spatially averaged value for the mobility,  $\bar{\mu}_{pE}$ , can be employed.

Since in modern transistors the emitter junction is very shallow a first-order approximation is to neglect the recombination in the mono-silicon layer. Thus, the same derivation as for the low injection transfer current in the base can be performed. Applying (3.50) with the assumptions made above to the emitter volume region leads to

$$I_{jBEi} = A_E \frac{q\bar{\mu}_{pE}V_T}{w_E} [\Delta p(w_E) - \Delta p(0)] \quad (3.95)$$

with  $\Delta p = p - p_0$  as the excess hole density w.r.t. the equilibrium density. The corresponding low-injection boundary condition at the BE junction reads

$$\Delta p(w_E) = p_0(w_E) \left[ \exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right] \quad (3.96)$$

with the equilibrium carrier density

$$p_0(w_E) = \frac{n_{iE}^2(w_E)}{N_E(w_E)}. \quad (3.97)$$

The boundary condition at  $x = 0$  depends on the physical effects occurring at the interface. Classical theory assumes  $\Delta p(0) = 0$ , which is not realistic for modern technologies with "semi"-transparent mono-silicon emitters. Thus, a more accurate approach is to assume an effective contact recomb-

nation velocity  $v_{Epm}$  that defines the hole density at  $x = 0$  through the current leaving the interface, i.e.

$$\Delta p(0) = \frac{-J_p(0)}{qv_{Epm}} = \frac{I_{jBEi}}{qA_E v_{Epm}}, \quad (3.98)$$

which is consistent with the assumption of negligible volume recombination within  $[0, w_E]$ . Inserting both boundary conditions in (3.95) yields for the back-injection current

$$I_{jBEi} = qA_E \frac{\bar{\mu}_{pE} V_T n_{iE}^2(w_E)}{w_E N_E(w_E)} \frac{\exp\left(\frac{V_{BE}}{V_T}\right) - 1}{1 + \frac{\bar{\mu}_{pE}(V_T/w_E)}{v_{Epm}}}. \quad (3.99)$$

The ratio of the diffusion velocity  $\bar{\mu}_{pE}(V_T/w_E)$  to the contact recombination velocity  $v_{Epm}$  determines the mechanism by which the magnitude of the back-injection current is controlled.

So far, the interface recombination velocity represents the various physical effects occurring at the interface and in the poly-silicon region. A comprehensive analysis of the hole current in the emitter, including a multi-grain poly-silicon region and corresponding interfaces, was given in [23]. As it turns out, the interface recombination velocity can be expressed as

$$v_{Epm} = v_{pmI} + \frac{1}{\frac{-1}{v_{TI}} + \frac{1}{v_{pmI} + v_{Gk}}} \quad (3.100)$$

where  $v_{TI}$  is a tunneling velocity, and  $v_{Gk}$  is the effective recombination velocity at the  $k$ th grain boundary (to be calculated recursively). Due to the strong dependence on processing conditions of the properties in the poly-silicon region and at the interface it does not seem to make sense from a practical application point of view to develop a more sophisticated model than (3.99).

For practical applications the results obtained above can be summarized for the quasi-static emitter back-injection current as

$$I_{BEi} = I_{BEiS} \left[ \exp\left(\frac{V_{B'E'}}{m_{BEi} V_T}\right) - 1 \right], \quad (3.101)$$

in which the saturation current  $I_{BEiS}$  contains all physical and structural parameters of (3.99).  $m_{BEi}$  is the emission (or ideality) coefficient with a typical range of 1.001...1.1. The deviation from 1.0 is caused in this component by a (small) change in the width  $w_E$  of the neutral emitter (cf. (3.99)), possible volume recombination, bias dependent tunneling, etc.. As Fig. 3.13 shows, (3.101) is an excellent approximation over the bias region of interest for circuit design beyond  $V_{B'E'} = 0.6\text{V}$  and even well beyond peak  $f_T$  where the error reaches just about 2%. Down to low injection, the error increases due to SRH recombination, which is not included in (3.101) and will be discussed in the next paragraph. The same result is obtained for a wide range of voltages  $V_{B'C'}$  between the hard-saturation and BC break-down region.

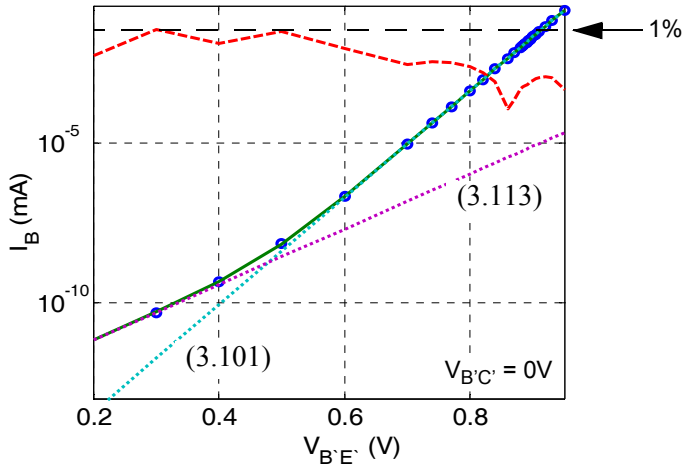


Fig. 3.13: Forward base current of the 1D BJT over a large bias range: comparison between analytical equations  $I_{BEi}$  and  $I_{REi}$  (solid lines with dotted line extensions) and device simulation results (symbols). The model parameters were calculated simply by using two bias points in each region resulting in  $m_{BEi} = 1.0022$ ,  $I_{BEiS} = 1.741 \cdot 10^{-20}\text{A}$  and  $m_{REi} = 1.9396$ ,  $I_{REiS} = 1.2745 \cdot 10^{-16}\text{A}$ . The dashed line shows the relative error.

### 3.3.3 Currents related to space-charge region recombination

In all useful production processes SCR recombination should be negligible within the relevant operating region for circuit design. Sometimes, however, transistors are operated at very low current densities where generally the relative contribution of SRH recombination in the BE SCR increases and causes the current gain to drop. In a real (2D or 3D) transistor there are also recombination centers at the surface and, due to higher doping, in the emitter periphery junction. At very high current densities, when the BC SCR vanishes and the collector region gets swamped with carriers, also some recombination might occur there. However, since usually the collector is only lightly to moderately doped, the diffusion length is larger than the collector width so that the collector recombination is negligible.

Below, a derivation for the 1D current component resulting from recombination in the BE SCR is given, which also reflects the different temperature dependence as compared to the back-injection component. The current flowing perpendicular through the junction area  $A_E$  is given by

$$I_{REi} = qA_E \int_0^{w_{BE}} R dx \quad (3.102)$$

with  $w_{BE}$  as the width of the BE depletion region. The well-known SRH recombination term reads [20,21,22]

$$R = \frac{pn - n_i^2}{\tau_p(n + n_i) + \tau_n(p + n_i)}, \quad (3.103)$$

in which for simplicity reasons the trap density is assumed to be located in the middle of the bandgap. Since the quasi-fermi potentials are constant throughout the depletion region, one can write the pn product as

$$pn = n_i^2 \exp\left(\frac{\Phi_p - \Phi_n}{V_T}\right) = n_i^2 \exp\left(\frac{V_{BE}}{V_T}\right). \quad (3.104)$$

Making the further simplifying assumption of equal electron and hole life time,  $\tau_p = \tau_n = \tau$ , with  $\tau$  averaged over the space-charge region, yields after inserting (3.104) into  $R$

$$R \cong \frac{n_i^2}{\tau(n+p+2n_i)} \left[ \exp\left(\frac{V_{B'E'}}{V_T}\right) - 1 \right]. \quad (3.105)$$

For the integration in (3.102), this is still a too complicated function of  $x$ . Hence, the effective intrinsic carrier density is taken as a spatially independent average value  $\bar{n}_i$ . Such averaging is reasonable due to the narrow width of the BE SCR. Furthermore,  $R$  is replaced by its maximum value, which occurs if  $(n+p)$  reaches a minimum. This happens under the condition

$$p = n = \bar{n}_i \exp\left(\frac{V_{B'E'}}{2V_T}\right), \quad (3.106)$$

where (3.104) was used. Inserting the result into (3.105) gives

$$R_{max} \cong \frac{\bar{n}_i^2}{2\tau\bar{n}_i \left[ \exp\left(\frac{V_{B'E'}}{2V_T}\right) + 1 \right]} \left[ \exp\left(\frac{V_{B'E'}}{V_T}\right) - 1 \right], \quad (3.107)$$

which does not depend on  $x$  anymore. Performing the integration in (3.102) then yields

$$I_{REi} = \frac{qA_E\bar{n}_i w_{BE}(V_{B'E'})}{2\tau \left[ \exp\left(\frac{V_{B'E'}}{2V_T}\right) + 1 \right]} \left[ \exp\left(\frac{V_{B'E'}}{V_T}\right) - 1 \right] \quad (3.108)$$

Since for any reasonably forward biased junction,  $\exp(V_{B'E'}/(2V_T)) \gg 1$ , the voltage dependence of the exponential terms can be combined as

$$\frac{\left[ \exp\left(\frac{V_{B'E'}}{V_T}\right) - 1 \right]}{\left[ \exp\left(\frac{V_{B'E'}}{2V_T}\right) + 1 \right]} = \exp\left(\frac{V_{B'E'}}{2V_T}\right) - 1, \quad (3.109)$$

which still gives the correct non-zero value at reverse bias. The resulting recombination current then reads

$$I_{REi} = I_{REiS} \left[ \exp\left(\frac{V_{B'E'}}{2V_T}\right) - 1 \right] \quad (3.110)$$

with the saturation current

$$I_{REiS} = qA_E w_{BE} (V_{B'E'}) \frac{\bar{n}_i}{2\tau}. \quad (3.111)$$

It still depends slightly on the bias voltage via the bias dependence of both the SCR and the (average) values of  $\tau$  and  $\bar{n}_i$ . The latter depend on the SCR width due to their doping dependence. In a non-symmetric pn junction, the total doping density decreases in direction of the lower doped region, into which the SCR mostly extends. Therefore, an increase of  $V_{B'E'}$  leads to a decrease of  $w_{BE}$  and, hence, to a decrease of  $\tau$  and increase of  $\bar{n}_i$  due to the larger total doping density within the remaining SCR. As a consequence, the bias dependence of  $w_{BE}$ ,  $\tau$  and  $\bar{n}_i$  partially compensate each other. This seems also to be confirmed by measurements and device simulations that have shown a negligible bias dependence of the recombination saturation current. Hence, the saturation current is generally considered a model parameter and can be written as

$$I_{REiS} \cong qA_E w_{BE0} \frac{\bar{n}_i}{2\tau_{eq}}, \quad (3.112)$$

with  $w_{BE0}$  and  $\tau_{eq}$  taken at equilibrium. To compensate for some of the simplifications and to provide modeling flexibility, an "emission" coefficient  $m_{REi}$  is introduced (with default value  $m_{REi} = 2$ ). This gives as final expression for the recombination current component

$$I_{REi} = I_{REiS} \left[ \exp\left(\frac{V_{B'E'}}{m_{REi} V_T}\right) - 1 \right]. \quad (3.113)$$

Figure 3.13 shows that the above approximation agrees very well with device simulation results at very low voltages ( $V_{B'E'} = 0.2 \dots 0.5V$ ). Towards  $V_{B'E'} = 0.2V$  the factor  $m_{REi}$  indeed approaches 2. The solid line in Fig. 3.13 corresponds to the sum  $i_{BE} = i_{jBEi} + i_{REi}$  of the two base current components. Through numerical fitting procedures, an even better and more balanced (over the bias region) agreement can be obtained.

### 3.3.4 Avalanche current from base-collector breakdown

At large BC reverse voltages carriers in the BC SCR can gain sufficient energy so that a collision with a valence band electron can cause the latter to become “free” and move up into the conduction band, while at the same time also a hole is generated through the ionization of the corresponding lattice atom. This process is called impact ionization. Once generated, the electron and hole will move in different directions according to the electric field, as shown in Fig. 3.14. Further ionizing collisions of the newly generated carriers are called secondary ionization and act like an avalanche. The generation of new carriers leads to an increase of the carrier current density through the SCR towards the ends. For a detailed description of the physics of the impact ionization process the reader is referred to, e.g., [2, 24, 25].

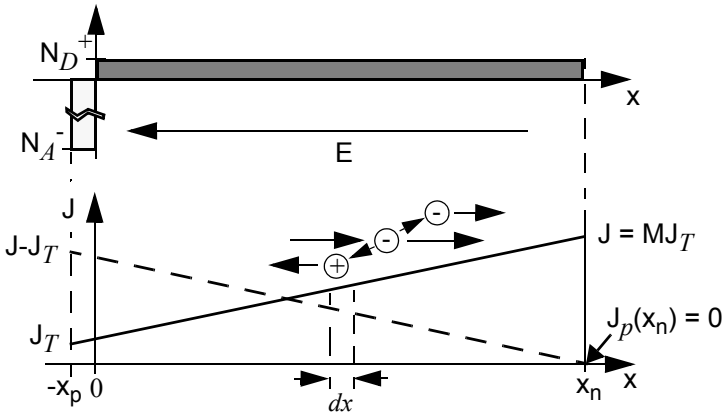


Fig. 3.14: Illustration of impact ionization in the BC SCR. The latter is defined by the edges  $-x_p$  and  $x_n$ . Upper figure: space charge and field direction. Lower figure: carrier flow direction after impact ionization and current density.

The multiplication factor resulting from the carrier avalanche is defined for each type of carriers as the ratio of the current density leaving the impact ionization region to the current density entering that region. For the electron current density that is of interest here,

$$M_n = \frac{J_n(x_n)}{J_n(-x_p)} = \frac{J}{J_T}. \quad (3.114)$$

If  $\alpha_n$  ( $\alpha_p$ ) is the probability with which *one* new electron (hole) is generated per spatial interval  $dx$ , then  $\alpha_n dx \cdot n$  ( $\alpha_p dx \cdot p$ ) is the absolute number of generated electrons (holes) in  $dx$ . Since the electrons (holes) move through the SCR with their average velocity  $v_{sn}$  ( $v_{sp}$ ), the increase of the (electrical) electron current density over the distance  $dx$  is

$$dJ_n = (-q)v_{sn}(\alpha_n dx \cdot n) + (q)v_{sp}(\alpha_p dx \cdot p). \quad (3.115)$$

With  $J_n(x) = -qv_{sn}n$  and  $J_p(x) = qv_{sp}p$  as well as assuming  $\alpha_n = \alpha_p = \alpha$  for simplification one obtains

$$dJ_n = \alpha(J_n(x) + J_p(x))dx = \alpha J dx. \quad (3.116)$$

Since the total current  $J$  must be spatially independent integration yields

$$J_n(x_n) - J_n(-x_p) = J \int_{-x_p}^{x_n} \alpha dx. \quad (3.117)$$

Inserting  $J_n(-x_p) = J_n(x_n)/M$  from (3.114) and  $J_n(x_n) = J$  according to Fig. 3.14 leads to

$$1 - \frac{1}{M} = \int_{-x_p}^{x_n} \alpha dx. \quad (3.118)$$

For *weak* avalanche, the ionization integral is quite small allowing to write

$$M \approx 1 + \int_{-x_p}^{x_n} \alpha dx. \quad (3.119)$$

Note, that even for weak avalanche due to the usually large current gain the hole current density can increase dramatically while an increase in electron current density is hardly visible.

Classical theory assumes an instant acceleration of carriers with the *local* driving force  $E(x)$ , which corresponds to a field. This assumption is justified for sufficiently large SCR widths. Experimentally, the ionization rate was found to be a function of the magnitude of the local field [25,26,27],

$$\alpha(E(x)) = a \exp(-b/|E(x)|), \quad (3.120)$$

with  $a$  and  $b$  as material and temperature dependent coefficients. As driving force, generally the gradient of the quasi-Fermi potential is used. Note that for electrons and holes the values of  $a$  and  $b$  are somewhat different. Also,  $a \sim |E|$  according to [26]; however, this dependence is small compared to the exponential one.



A relation between  $M$  and the desired avalanche current can be derived by noting that according to Fig. 3.14,  $J_n(-x_p) = J_T = -I_T/A_E$  and with (3.114)

$$I_C = -J_n(x_n)A_E = MI_T. \quad (3.121)$$

Since also  $I_C = I_{AVL} + I_T$ , the avalanche current is given by

$$I_{AVL} = (M - 1)I_T. \quad (3.122)$$

For determining  $M$  from either (3.118) or (3.119) the integral

$$\int_{-x_p}^{x_n} \alpha dx = \int_{-x_p}^{x_n} a \exp\left(-\frac{b}{|E(x)|}\right) dx \quad (3.123)$$

needs to be evaluated, where (3.120) was used for the ionization rate. The analytical evaluation requires an assumption about the spatial dependence of the electric field. Since the width of the BC SCR is given mostly by the extension into the collector, the electric field distribution in the collector at low current densities can be approximated by

$$|E(x)| = E_{jc} \left(1 - \frac{x}{w_{BC}}\right), \quad (3.124)$$

with  $E_{jc}$  as the (absolute) peak value of the field at the BC junction ( $x = 0$ ). Due to the exponential dependence of the ionization coefficients on the local electric field, the ionization process occurs mostly in the vicinity of the maximum field which is located at the junction. As a consequence, one can write (e.g. [28])

$$\frac{1}{|E(x)|} \approx \frac{1}{E_{jc}} \left(1 + \frac{x}{w_{BC}}\right). \quad (3.125)$$

Insertion of (3.125) into (3.123) yields after integration over  $[0, w_{BC}]$

$$M = \left(1 - \frac{aE_{jc}w_{BC}}{-b} \exp\left(-\frac{b}{E_{jc}}\right) \left[\exp\left(-\frac{b}{E_{jc}}\right) - 1\right]\right)^{-1}. \quad (3.126)$$

Since  $b$  is in the order of 1MV, the exponential term in the brackets is negligible in the application relevant BC voltage range allowing to write

$$M = \left[1 - \frac{aE_{jc}w_{BC}}{b} \exp\left(-\frac{b}{E_{jc}}\right)\right]^{-1}. \quad (3.127)$$

For weak avalanche, one obtains (in accordance with (3.119))

$$M \approx 1 + \frac{a_n E_{jc} w_{BC}}{b_n} \exp\left(-\frac{b_n}{E_{jc}}\right). \quad (3.128)$$

Inserting above multiplication factors into (3.122) gives the desired avalanche current as function of physical parameters and bias.

Figure 3.15 shows a comparison of the multiplication factor, determined and calculated in different ways. The calculation of the ionization integral serves as reference. Inserting the electric field (i.e.  $|\text{grad } \varphi_n|$ ) in (3.120) allows to calculate  $M$  directly from the internal results of device simulation using (3.118). This equation is the simple form of the ionization integral in which the difference between electron and hole ionization coefficient is neglected. (For the general integral expressions see, e.g., [25]). As shown in Fig. 3.15, this difference reduces  $M$  somewhat. In general,  $M$  is determined from measurements using the appropriate terminal currents. In case of a bipolar transistor,

$$I_{AVL} = I_{B0} - I_B(V_{C'B'}), \quad (3.129)$$

where  $I_{B0}$  is the current without avalanche effect at sufficiently low  $V_{C'B'}$ . With  $I_C = I_{AVL} + I_T$  and (3.122) the experimentally determined multiplication factor (for an npn transistor) is given by

$$M_n = 1 + \frac{I_{AVL}}{I_C - I_{AVL}}, \quad (3.130)$$

which is also inserted in Fig. 3.15. The respective results are very close to those from the general ionization integral, which provides a good feeling for the accuracy of this measurement method.

Finally, Fig. 3.15 also contains the results obtained from the compact equations (3.127) and (3.128). Here, the electric field  $E_{jc}$  was taken directly from device simulation, and  $w_{BC} = \varepsilon/C_{jCi}$  (e.g. [29]) was used with  $C_{jCi}$  calculated from small-signal analysis. Notice that the model for weak avalanche is very close to the reference while (3.127) exhibits somewhat higher deviations and is closer to the simple ionization integral results. Nevertheless, the compact expressions appear to be quite suitable for accurately approximating the multiplication factor and, thus, the avalanche current.

For a constant collector doping profile, classical theory allows to link field and SCR width:

$$E_{jc} = 2 \frac{V_{DCi} - V_{B'C}}{w_{BC}}. \quad (3.131)$$

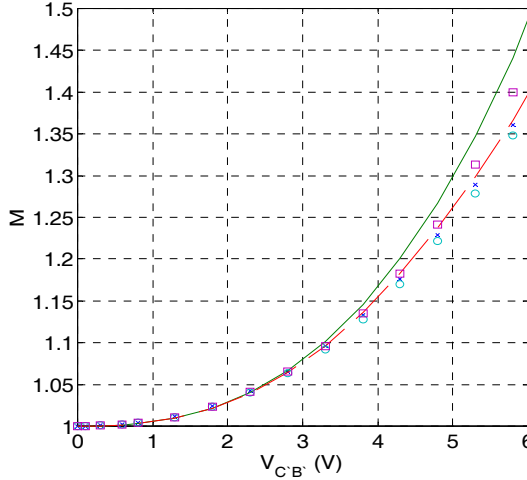


Fig. 3.15: Avalanche multiplication factor vs. reverse BC voltage for the BJT: comparison between (3.127) (solid line), (3.128) (dashed line), (3.130) (x),  $M$  from the simple ionization integral (3.119) (squares), and from (3.118) (circles).  $V_{BE} = 0.7V$ .

Furthermore, the SCR width can be expressed by the BC depletion capacitance,

$$w_{BC} = \varepsilon / C_{jCi} \quad (3.132)$$

which is always a continuously differentiable function of bias in a compact model. Inserting the last two equations back into (3.128) yields a compact equation for the avalanche current

$$I_{AVL} = I_T \frac{2a_n(V_{DCi} - V_{B'C})}{b_n} \exp\left(-\frac{b_n \varepsilon}{2C_{jCi}(V_{DCi} - V_{B'C})}\right). \quad (3.133)$$

The potential discontinuity at  $V_{B'C} = V_{DCi}$  can be eliminated by, e.g., also replacing the expression  $(V_{DCi} - V_{B'C})$  with the depletion capacitance. The choice of the respective smoothing function, which is model depen-

dent, is uncritical though since the discontinuity occurs at high forward bias across the BC junction.

The derived equation describes the weak avalanche effect at sufficiently low current densities, when the electric field peaks at the BC junction and is independent of current. The dependence on  $I_T$  may be included approximately through a current dependent depletion capacitance. Furthermore, non-local transport effects have been ignored so that the obtained relation can only be applied to transistors with sufficiently large space charge regions. In other words, the calculated local breakdown current is larger than the actual (non-local) current and, thus, the breakdown voltage calculated from local theory is lower than the non-local voltage. As long as the local field approximation holds, the weak avalanche current derived can be used to describe fairly accurately the onset of the avalanche effect as well as the respective breakdown voltages  $BV_{CEO}$  and  $BV_{CBO}$ . The latter is often directly related to  $M$  by the empirical expression [30]

$$M = \frac{1}{1 - (-V_{B'C}/BV_{CBO})^k} \quad (3.134)$$

with  $k = 3 \dots 5$  as a material and doping dependent factor.

A relation between  $BV_{CEO}$  and  $BV_{CBO}$  can be obtained as follows. Inserting  $M$  and  $V_{B'C}$  at the condition  $I_B = 0$  into (3.134) gives

$$M(BV_{CEO}) = \frac{1}{1 - ((BV_{CEO} - V_{B'E})/BV_{CBO})^k} = 1 + \frac{I_{AVL}}{I_T}$$

with  $V_{B'E}(I_T) = \text{const}$ . Noticing that under this condition  $I_{AVL}$  in (3.129) equals  $I_{B0} = I_T/B_f$  with  $B_f$  as forward DC common-emitter current gain, one obtains for the measured open-base breakdown voltage

$$BV_{CEO} = BV_{CBO} \left( \frac{1}{1 + B_f} \right)^{1/k} - V_{B'E} \left( \frac{I_T}{B_f} \right). \quad (3.135)$$

Letting as per definition  $V_{B'E} \rightarrow 0$  yields the often found expression

$$BV_{CEO} = BV_{CBO} \left( \frac{1}{1 + B_f} \right)^{1/k}. \quad (3.136)$$

### 3.4 Depletion (junction) charges and capacitances

The depletion of carriers around the junctions leads to a depletion (or often called junction) charge and capacitance. Their effect on electrical transistor and circuit characteristics is mostly visible at low-current densities during large-signal switching, high-frequency small-signal, and even DC operation. For instance, the depletion charges also determine the Early-effect, while the depletion capacitances are a limiting factor for device and circuit performance in low-power applications. In high-frequency analog amplifiers, the BC depletion capacitance not only determines the voltage gain (Miller-effect) but also can limit the achievable linearity. These are just a few examples showing the importance of accurate modeling of depletion capacitances.

In the next sections, the results of the classical theory for depletion capacitance and charge are reviewed first and evaluated based on device simulation. This will reveal the major shortcomings of these equations for practical applications. Possible extensions will then be presented in subsequent sections.

#### 3.4.1 Classical approach

The classical derivation of analytical expressions for the depletion charge and capacitance usually starts with a doping profile that contains an *abrupt* transition from, e.g., the donor concentration  $N_D$  to the acceptor concentration  $N_A$ . Under the assumption of negligible mobile carriers ("depletion approximation") and spatially independent permittivity within the SCR, Poisson's equation is solved, leading to the SCR width (e.g. [1,25]):

$$w(V) = \sqrt{\frac{2\epsilon}{q} \left( \frac{N_A + N_D}{N_A N_D} \right) (V_D - V)} = w_0 \sqrt{1 - \frac{V}{V_D}}. \quad (3.137)$$

Here,  $V$  is the forward voltage across the SCR, and  $V_D$  is the built-in voltage, which represents the potential barrier in equilibrium and is discussed later, and  $w_0$  is the zero-bias SCR width. Under above assumptions and realizing that the SCR is charged only at its edges (cf. Fig. 2.9) with distance  $w$  apart, the corresponding capacitance can be simply calculated as

$$C_j(V) = A_E \frac{\varepsilon}{w(V)}. \quad (3.138)$$

Inserting  $w$  from (3.137) this is generally written in the form

$$C_j = \frac{C_{j0}}{(1 - V/V_D)^z}, \quad (3.139)$$

with the zero-bias capacitance

$$C_{j0} = A_E \sqrt{\frac{\varepsilon q}{2V_D} \left( \frac{N_A N_D}{N_A + N_D} \right)} \quad (3.140)$$

and the exponent factor  $z = 1/2$ . For a linearly graded profile  $z$  equals  $1/3$  and also the expression for  $C_{j0}$  changes. For practical applications with realistic doping profiles  $z$  is generally in between the above values, and  $(C_{j0}, V_D, z)$  are used as model parameters that are adjusted to measured data for  $C_j(V)$ .

Integration of (3.139) from zero bias to the voltage  $V$  gives the corresponding depletion charge

$$Q_j(V) = \frac{C_{j0} V_D}{1 - z} \left[ 1 - \left( 1 - \frac{V}{V_D} \right)^{(1-z)} \right]. \quad (3.141)$$

This equation together with (3.139) is the basis for depletion charge and capacitance formulations in all compact models.

Figure 3.16a,b shows the voltage dependence of the BC and BE depletion capacitance of the 1D BJT. (Note that the collector doping was reduced to  $2 \cdot 10^{16} \text{cm}^{-3}$  which is similar to power transistors in the same process technology.) As can be seen, the analytical expression (3.139) agrees quite well for not too high forward and reverse bias. The model parameters were determined with a simple method; i.e. numerical fitting may yield a somewhat improved result, but will not eliminate the deviations at higher forward and reverse bias. The deviations of the classical expression and their causes will be discussed in subsequent sections.

The built-in voltage that appears in the derivation of (3.139) requires a closer inspection. Its derivation starts from the 1D equilibrium transport equation of holes or electrons ( $c = p, n$ )

$$q\mu_c V_T \frac{dc}{dx} = \pm q\mu_c c \left( -\frac{d\psi}{dx} \pm \frac{V_T}{n_{ie}} \frac{dn_{ie}}{dx} \right), \quad (3.142)$$

where the upper (lower) sign is for holes (electrons). The  $dn_{ie}/dx$  term includes a spatial bandgap variation caused, e.g., by high-doping effects and, in HBTs, also intentional bandgap grading. With  $n_{ie} = n_{i0} \exp(\Delta V_g / (2V_T))$  from (2.7) the derivative term can also be written as

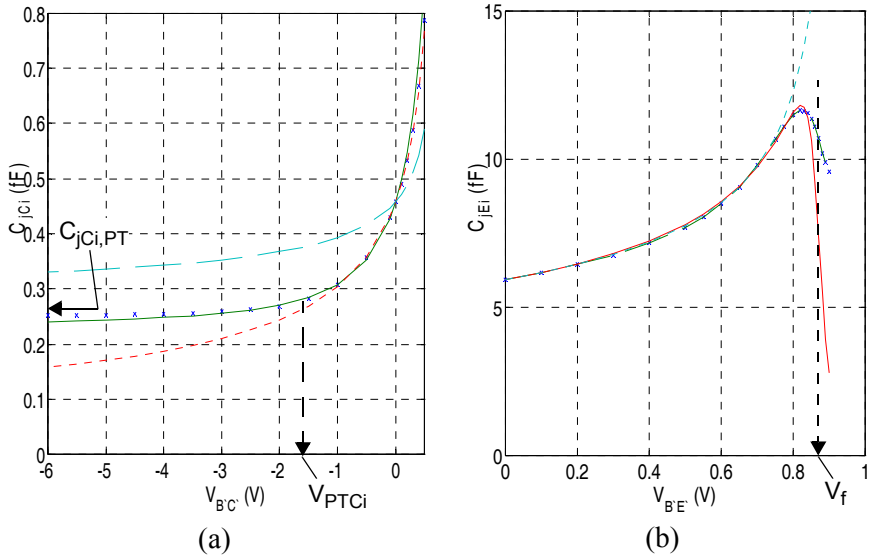


Fig. 3.16: Depletion capacitances vs. voltage for the BJT with  $N_{Ci} = 2 \cdot 10^{16} \text{ cm}^{-3}$ . Comparison between device simulation (symbols), classical approximation (3.139) (dotted lines), and extended analytical approximations (3.163) and (3.158) (solid lines). (a)  $C_{jCi}$  vs. reverse voltage  $V_{BC'}$ , with adjusted model parameter set  $(V_{DCi}, z_{Ci}, C_{jCi, PT}) = (1.09 \text{ V}, 1.646, 0.23 \text{ fF})$  and physical model parameter set  $(0.787 \text{ V}, 0.5, 0.264 \text{ fF})$  for (3.163) as well as with  $(V_{DCi}, z_{Ci}) = (0.787 \text{ V}, 0.5)$  for the classical eq. (3.139). (b)  $C_{jEi}$  vs. forward voltage  $V_{BE'}$ , with adjusted model parameter set  $(V_{DEi}, z_{Ei}, V_f) = (0.91 \text{ V}, 0.337, 0.868 \text{ V})$  for (3.158) and with  $(V_{DEi}, z_{Ei}) = (0.9 \text{ V}, 0.333)$  for (3.139); the factor  $a_{fi}$  was set to 1.92.

$$\frac{dn_{ie}}{dx} = n_{i0} \exp\left(\frac{\Delta V_g}{2V_T}\right) \frac{1}{2V_T} \frac{d\Delta V_g}{dx} = \frac{n_{ie}}{2V_T} \frac{d\Delta V_g}{dx}. \quad (3.143)$$

Insertion into (3.142) yields after cancelling variables that occur on both sides of the equations and rearranging terms gives

$$\frac{1}{c} \frac{dc}{dx} = \pm \frac{1}{V_T} \left( -\frac{d\psi}{dx} \pm \frac{1}{2} \frac{d\Delta V_g}{dx} \right). \quad (3.144)$$

Integration over  $dx$  between the edges of the SCR,  $-x_p$  and  $x_n$ , gives for the hole equation

$$V_T \ln \left( \frac{p(x_n)}{p(-x_p)} \right) = [\psi(-x_p) - \psi(x_n)] + \frac{1}{2} (\Delta V_g(x_n) - \Delta V_g(-x_p)). \quad (3.145)$$

Similarly, one obtains for the electron equation

$$V_T \ln \left( \frac{n(x_n)}{n(-x_p)} \right) = \psi(x_n) - \psi(-x_p) + \frac{1}{2} (\Delta V_g(x_n) - \Delta V_g(-x_p)). \quad (3.146)$$

Defining the built-in voltage as

$$V_D = \psi(x_n) - \psi(-x_p) > 0, \quad (3.147)$$

the bandgap-difference across the junction as

$$\Delta V_{gpn} = \Delta V_g(x_n) - \Delta V_g(-x_p), \quad (3.148)$$

and the carrier densities at the SCR edges as (index 0 for equilibrium)

$$\begin{aligned} p_{n0} &= p(x_n), \quad p_{p0} = p(-x_p), \quad n_{p0} = n(-x_p), \\ n_{n0} &= n(x_n), \end{aligned} \quad (3.149)$$

gives for  $V_D$  from (3.145) and (3.146), respectively,

$$V_D = V_T \ln \left( \frac{p_{p0}}{p_{n0}} \right) + \frac{\Delta V_{gpn}}{2} \text{ and } V_D = V_T \ln \left( \frac{n_{n0}}{n_{p0}} \right) - \frac{\Delta V_{gpn}}{2} \quad (3.150)$$

The equilibrium minority carrier densities are given by the pn product

$$p_{n0} = \frac{n_{i0}^2}{n_{n0}} \exp \left( \frac{\Delta V_g(x_n)}{V_T} \right) \text{ and } n_{p0} = \frac{n_{i0}^2}{p_{p0}} \exp \left( \frac{\Delta V_g(-x_p)}{V_T} \right). \quad (3.151)$$

Inserting these into (3.150) and (3.150) yields in both cases

$$V_D = V_T \ln \left( \frac{p_{p0} n_{n0}}{n_{i0}^2} \right) - \Delta V_{g\Sigma}. \quad (3.152)$$



The bandgap variation occurs as the sum of the variations on either side of the junction and can be lumped into the term

$$\Delta V_{g\Sigma} = \frac{\Delta V_g(x_n) + \Delta V_g(-x_p)}{2}. \quad (3.153)$$

If  $\Delta V_g > 0$ , which corresponds to a bandgap reduction,  $V_D$  is lowered, too.

Note, that the above result is valid up to moderate bias in forward and, depending on the SCR width, also reverse bias. For the case of not too high or too low temperatures and negligible bandgap variation across the depletion region, one can replace the majority carrier densities at the edges of the equilibrium SCR width by the doping densities to obtain the classical result

$$V_D = V_T \ln \frac{N_D N_A}{n_{i0}^2}. \quad (3.154)$$

In most semiconductor devices the doping concentration on one side is much lower than the concentration on the other side. For such highly non-symmetric pn-junctions the capacitance expression (3.140) simplifies to

$$C_{j0} = A_E \sqrt{\frac{\epsilon q N}{2 V_D}}. \quad (3.155)$$

with  $N$  as the concentration of the lower doped side. This relation together with (3.152) links device structure and physics information to important electrical quantities with quite reasonable accuracy that is acceptable for many practical applications, including statistical modeling. There is no such relation known for the exponent factor.

### 3.4.2 High forward bias operation

Under high-forward bias the assumption of negligible mobile carriers within the SCR breaks down. As a consequence, the results from classical theory cannot be used anymore. This already became evident from Fig. 3.16. Besides an increasing error compared to the actual depletion capacitance, the classical approximation also causes a numerical overflow at  $V = V_D$  that needs to be avoided by a proper modification.

Physically, an increasing forward bias across the junction leads to a decrease of the SCR width and an increase of the mobile carrier densities therein. A qualitative explanation for the observed voltage dependence can be easily obtained by considering an abrupt doping profile. At some bias point the space-charge density around the junction will be completely compensated and will have disappeared. This is equivalent to a "saturation" of the space-charge  $Q_j$  from (3.141), which now equals the value  $-qA_ENw_0$ . Therefore, the derivative of the actual space-charge  $Q_j$ , i.e. the depletion capacitance, has to approach zero towards high forward bias. This indicates that the depletion capacitance must be limited to some maximum value before it drops to zero, which in fact can be observed from Fig. 3.16.

The behavior of the SCR width and junction capacitance at high forward bias has been the subject of numerous investigations (e.g. [34,31,32,33]). Starting from Poisson's equation, in all cases eventually a numerical solution of the derived relation becomes necessary. However, the probably most important insight into the direction of modifying the classic relation was given in [31, 32] in which it was realized that at a high forward bias the built-in voltage changes with applied voltage. Along these lines several modifications of the expression  $(V - V_D)$  were suggested in literature. Most of them target directly the depletion capacitance without considering the compact model related requirement for a smooth and explicit charge expression. Among those suitable for compact modeling are a simple linearization (cf. SPICE GP model, e.g. [17]) and more sophisticated smoothing functions such as those in advanced compact models. A practically and numerically suitable approach is to include the bias dependence of  $V_D$  in the applied voltage and to replace latter by the auxiliary voltage

$$V_j = V_f - V_T \frac{x + \sqrt{x^2 + a_{ff}}}{2} < V_f \quad (3.156)$$

using a *hyperbolic* smoothing expression with the argument

$$x = \frac{V_f - V}{V_T}. \quad (3.157)$$

The voltage  $V_f$  represents the upper limit of  $V_j$  and is considered as a fitting parameter. The value of  $a_{ff}$  can be adjusted to control the “smoothness” of the approximation. The function (3.156) yields  $V_j = V$  for  $V < V_f$  and approaches  $V_j = V_f$  for  $V > V_f$ , limiting the charge value to its maximum.

Employing  $V_j$  in the charge expression leads to a constant charge value at high forward bias. The capacitance is calculated from the derivative of the charge yielding

$$C_j = \frac{C_{j0}}{(1 - V_j/V_D)^z} \cdot \frac{dV_j}{dV} \quad (3.158)$$

with the derivative term

$$\frac{dV_j}{dV} = \frac{x + \sqrt{x^2 + a_{ff}}}{2\sqrt{x^2 + a_{ff}}} \quad (3.159)$$

Figure 3.16 also shows a comparison of above equation with device simulation results. In order to optimize its accuracy especially close to the peak region,  $V_D$  turned out to be slightly different from the value obtained for the classical equation (from the lower voltage range). It is important though to keep in mind that the accuracy of the approximation beyond the peak is of little importance for practical applications due to the much larger diffusion capacitance (cf. Fig. 2.11) in parallel. Notice that the ratio of peak to zero-bias value does not vary much for different doping profiles.

### 3.4.3 Punch-through case

According to Fig. 3.16, the classical capacitance expression also deviates from the actual behavior at sufficiently high reverse bias. Especially for larger voltages this deviation can become quite significant, and circuit simulation may give too optimistic results if this effect is not considered. The reason for the deviation is a fully depleted epi-collector region between base and buried layer. This case is called *collector punch-through*. The SCR width increase with reverse voltage is then much smaller due to the much higher doping of the buried layer compared to the epi-collector. Once the voltage is low enough so that the BC SCR is not fully depleted anymore the capacitance follows the classical relation.

Since in punch-through the SCR width is more or less fixed by the distance  $w_{Ci}$  between base and buried layer, the punch-through capacitance is approximately given by

$$C_{jCi, PT} = \frac{\varepsilon}{w_{Ci}}. \quad (3.160)$$

Assuming a constant doping concentration  $N_{Ci}$  and negligible mobile carriers, the applied (reverse) voltage at which punch-through occurs can be calculated from (3.137) by setting  $N_D = N_{Ci} \ll N_A = N_B$ :

$$w_{Ci} = w(V_{PTCi}) = \sqrt{\frac{2\varepsilon}{qN_{Ci}}(V_{DCi} + V_{jPCi})}. \quad (3.161)$$

Solving for the *applied* BC punch-through voltage gives

$$V_{jPCi} = \frac{qN_{Ci}}{2\varepsilon} w_{Ci}^2 - V_{DCi} = V_{PT} - V_{DCi}. \quad (3.162)$$

with the collector punch-through voltage  $V_{PT}$  from (3.8), which obviously represents the *internal* potential difference ( $V_{DCi} + V_{jPCi}$ ).

The most simple way to combine the punch-through capacitance (3.160) with the classical voltage dependent expression would be a simple case distinction. However, the resulting expression is not continuously differentiable which can lead to problems in certain compact model formulations as well as for distortion calculations. A simple smooth expression is

$$C_{jCi} = \frac{(C_{jCi0} - C_{jCi, PT})}{(1 - V_{B'C}/V_{DCi})^{z_{Ci}}} + C_{jCi, PT}, \quad (3.163)$$

which is compared to device simulation results in Fig. 3.16. Reasonable agreement is obtained after determining the model parameter set ( $V_{DCi}$ ,  $z_{Ci}$ ,  $C_{jCi, PT}$ ) with focus on the reverse bias region especially around the punch-through voltage. This agreement comes at the expense of (i) larger deviations already at low forward bias and (ii) significant deviations of the model parameters  $V_{DCi}$  and  $z_{Ci}$  from their physics-based values used for the classical expression. The latter can be understood by realizing that adding a constant capacitance simply “flattens” the curve, which in turn needs to be compensated for by mainly a larger value for  $z_{Ci}$ . Therefore, Fig. 3.16 also contains the result that is obtained with a physics-based values for the

parameter set ( $V_{DCi}$ ,  $z_{Ci}$ ,  $C_{jCi,PT}$ ). It is obvious that this is not an acceptable solution. It can also be noticed from Fig. 3.16 that the actual capacitance still slightly decreases towards large reverse and, in fact, falls below the (indicated) physics-based value for  $C_{jCi,PT}$ . This effect can be (relatively) more pronounced at higher collector doping and is caused by the continuing encroachment of the SCR into the higher doped regions of base and buried layer. Hence, in general the capacitance behavior is composed of two separate regions with different voltage dependence.

The charge corresponding to (3.163),

$$Q_{jCi} = \frac{(C_{jCi0} - C_{jCi,PT})V_{DCi}}{1 - z_{Ci}} \left[ 1 - \left( \frac{1 - V_{B'C'}}{V_{DCi}} \right)^{(1 - z_{Ci})} \right] + C_{jCi,PT}V_{B'C'}, \quad (3.164)$$

is for the classical part still given by the form of (3.141) with just a modified zero-bias value, while the punch-through capacitance (3.160) leads to an additional term. The deviation of above charge formulation from the one calculated from device simulation is very small. Hence, the modeling of large-signal transient behavior is less challenging as the accurate description of small-signal HF operation.

Note, that (3.163) and (3.164) still are numerically unstable at high forward bias and need to be modified accordingly. This eventually leads to more complicated equations like those employed in advanced compact models.

### 3.5 Mobile charge and storage times

The various contributions to the overall storage (or transit) time will be investigated and discussed in this section. It will be demonstrated that even the classical formulas are still useful estimates if they are properly used or modified. The discussion is generally divided into the cases of low, medium and high current densities as defined in section 3.1. The main focus is on the forward operation region, i.e.  $V_{C'E'} > 0$ , simply because it is of interest for the vast majority of applications. The resulting analytical expressions are the basis for equations used in different compact models.

### 3.5.1 Low injection

The designation of "low injection" is interchangeable with "low current densities". In this bias regime classical theory only considers the transit time in the neutral base region (e.g. [3, 1]). The well-known textbook expression for the base transit time at forward operation was derived from solving the transport differential equation in the same way as in section 3.3, but assuming an infinite velocity  $v_c$  of electrons in the BC space charge region. This latter assumption was dropped in [2], where also the delay time through the BC SCR was discussed. However, high-doping effects, the influence of the internal collector series resistance, as well as the storage time contributions from the neutral emitter,  $\tau_{pE}$ , and from the BE space charge region,  $\tau_{BE}$ , are neglected in classical theory. This is no longer possible in modern transistors as was shown in Fig. 2.12.

More sophisticated theoretical expressions for the base transit time taking into account spatial variations in doping, mobility and bandgap were published in, e.g., [35, 36]. In these cases a different derivation path as compared to classical theory was pursued which led to an *integral* and more general appearing representation of a storage time. Unfortunately, this minority storage time is always defined as the ratio of the *static* charge and current,  $\tau_m = Q_m/I_T$ , which is only equal to the small-signal storage time measured via the device terminals if  $Q_m$  is a linear function of current. As a consequence, the results are only applicable to the low injection region. Compact analytical expressions can be derived from such integral representations by making simplifying assumptions about the spatial dependence of doping, mobility and bandgap in much the same way as when solving a differential equation. In fact, when trying to apply the differential definition of storage times, which requires to include the bias dependence of the integration boundaries, integral representations are more difficult to deal with than differential equation solutions. So far it has been impossible to derive a "master" equation for the small-signal transit time similarly to that for the transfer current (cf. sec. 4.3). Therefore, the derivation of bias dependent minority charge components below is mostly based on the solution of a differential equation for carrier transport.

### 3.5.1.1 Neutral base region

Inserting  $f_z$  from (3.49) into (3.48) leads to

$$n_e - \frac{n_c}{f_z} = \frac{I_T w_B}{q A_E \bar{\mu}_{nB} V_T \zeta} \left[ 1 - \frac{1}{f_z} \right] \quad (3.165)$$

and permits to substitute the  $I_T$  related expression in (3.45) to obtain the spatial distribution of the electron density in general form as a function of the electron densities  $n_e$  and  $n_c$  at the boundaries of the neutral base:

$$n(x) = \frac{f_z - \exp\left(\zeta \frac{x - x_e}{w_B}\right)}{f_z - 1} n_e + \frac{\exp\left(\zeta \frac{x - x_e}{w_B}\right) - 1}{f_z - 1} n_c. \quad (3.166)$$

Integration of  $n(x)$  gives the minority charge stored in the neutral base,

$$Q_{nB} = \underbrace{q A_E w_B \frac{\zeta(f_z - 1) - 1}{\zeta(f_z - 1)}}_{\text{}} n_e + q A_E w_B \frac{f_z - 1 - \zeta}{f_z - 1} n_c. \quad (3.167)$$

In classical theory, this result is viewed as a superposition of two linearly independent contributions that are controlled by  $n_e$  and  $n_c$ , respectively. This approach leads to difficulties though at medium and high collector current densities once  $n_c$  becomes dependent on  $V_{B'E'}$  and, thus, the carrier densities are not linearly independent of each other anymore. In fact, even at low current densities  $n_c$  does depend on both  $V_{B'C'}$  and  $I_T$  due to the finite carrier velocity in the BC SCR.

From a measurement and high-speed application point of view it is advantageous to express  $Q_{nB}$  as a function of  $I_T$ . This is easy at low current densities since  $n_c$  is given by the finite drift velocity in the BC space charge region according to (3.2). Rearranging terms in (3.45) leads to the electron density as a function of  $I_T$ :

$$n(x) = \left[ \frac{-I_T}{q A_E v_c} + \frac{I_T w_B}{q A_E \bar{\mu}_{nB} V_T \zeta} \right] \left[ 1 - \exp\left(-\zeta \left(1 - \frac{x - x_e}{w_B}\right)\right) \right] + \frac{-I_T}{q A_E v_c}. \quad (3.168)$$

The spatial dependence of  $n(x)$  is shown in Fig. 3.17a for different values of the drift factor and both a finite and infinite drift velocity. The density is normalized to its injection density for a diffusion transistor with

infinite collector velocity,  $n_{e0} = I_T w_B / (q A_E \bar{\mu}_{nB} V_T)$ . Hence, the ratio of the diffusion velocity  $v_{diff} = \bar{\mu}_{nB} V_T / w_B$  to  $v_c$  was chosen as parameter. The electron density disappears at  $x = w_B$  for  $v_c \rightarrow \infty$ , i.e.  $v_{diff}/v_c = 0$ . The curves with  $\zeta = 0$  correspond to a diffusion transistor in which  $n$  decreases linearly from  $x_e$  to  $x_c$ , while  $\zeta = 8$  corresponds to a drift transistor in which the electron density is constant across almost the entire neutral base. Realistic values for  $\zeta$  in BJTs are typically between 1 and 4.

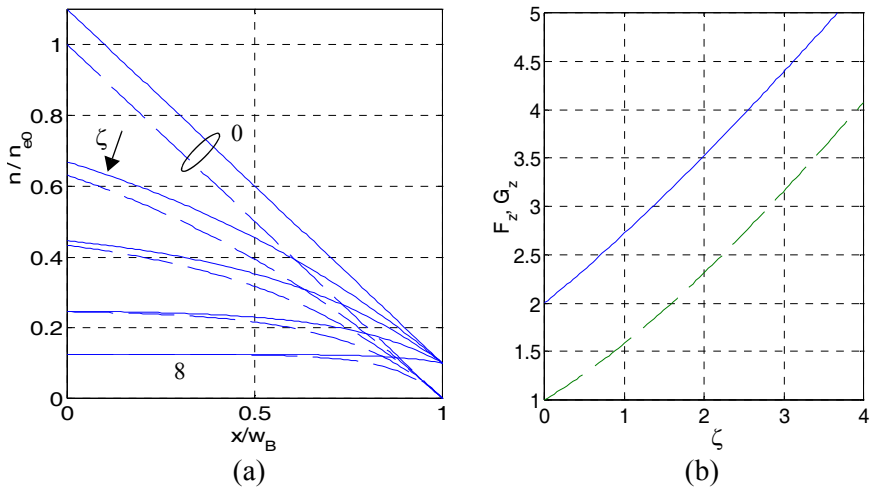


Fig. 3.17: (a) Normalized spatial dependence of electron density in the neutral base, calculated from (3.168) for the drift factors  $\zeta = 0, 1, 2, 4, 8$  with finite drift velocity  $v_{diff}/v_c = 0.1$  (solid lines) and infinite drift velocity  $v_{diff}/v_c = 0$  (dashed lines). (b) Base transport related drift functions  $F_\zeta$  (solid line) and  $G_\zeta$  (dashed line), respectively, from (3.172) and (3.175), respectively.

Integration of (3.168) over the neutral base region gives the desired stored charge as a function of  $I_T$ :

$$Q_{nB} = \left[ \frac{w_B^2}{\bar{\mu}_{nB} V_T} \frac{(\zeta - 1)f_\zeta + 1}{\zeta^2 f_\zeta} + \frac{w_B f_\zeta - 1}{v_c f_\zeta \zeta} \right] I_T. \quad (3.169)$$

Assuming that the current density is sufficiently low so that  $v_c$  and  $w_B$  do not depend on the transfer current, the storage time of the neutral base under forward operation mode reads



$$\tau_{Bf} = \frac{w_B^2}{\bar{\mu}_{nB} V_T} \frac{(\zeta - 1)f_\zeta + 1}{\zeta^2 f_\zeta} + \frac{w_B f_\zeta - 1}{v_c f_\zeta \zeta}. \quad (3.170)$$

The first term,

$$\tau_{Bfd} = \frac{w_B^2}{F_\zeta \bar{\mu}_{nB} V_T} \quad (3.171)$$

with the injection related drift function

$$F_\zeta = \frac{\zeta^2 f_\zeta}{(\zeta - 1)f_\zeta + 1}, \quad (3.172)$$

is known from the theory of drift transistors [37, 3, 1, 2]. For the special case of a diffusion transistor ( $\zeta = 0$ ) the factor  $F_\zeta$  approaches 2, and one obtains the well-known classical result

$$\tau_{Bfd, \zeta 0} = \tau_{Bfd}(\zeta \rightarrow 0) = \frac{w_B^2}{2\bar{\mu}_{nB} V_T}. \quad (3.173)$$

The second term in (3.170),

$$\tau_{Bfv} = \frac{w_B}{G_\zeta v_c} \quad (3.174)$$

with the collector exit related drift function

$$G_\zeta = \frac{f_\zeta \zeta}{f_\zeta - 1}, \quad (3.175)$$

represents the influence of the finite velocity in the BC space charge region which leads to a carrier jam in the base and limits the transport through the base. For a diffusion transistor the factor  $G_\zeta$  approaches 1. In classical theory  $\tau_{Bfv}$  is neglected, and the transit time  $\tau_f$  is set equal to  $\tau_{Bfd}$ .

Figure 3.17b shows the dependence of  $F_\zeta$  and  $G_\zeta$ , which correspond to the reciprocal normalized storage time components, as a function of the drift factor  $\zeta$ . At small fields these factors approach their limiting values for a diffusion transistor, while for large fields these factors are proportional to  $\zeta$ , so that both  $\tau_{Bfd}$  and  $\tau_{Bfv}$  become proportional to  $1/\zeta$ .

When calculating  $\tau_{Bf}$  in (3.170), it was assumed that the terms in the bracket of (3.169) do not depend on current. However, any transfer current variation is associated with a variation of the terminal voltage(s). Since the BE and BC SCR width depend on the applied voltages, so does the base width. Denoting the SCR width extension into the base on either side with  $w_{BE,b}$  and  $w_{BC,b}$ , respectively, the neutral base width is generally given by

$$w_B = w_{Bm} - w_{BE,b}(V_{B'E'}) - w_{BC,b}(V_{B'C'}, I_T) \quad (3.176)$$

with  $w_{Bm}$  as metallurgical base width. For the low injection region considered here the current dependence of  $w_{BC,b}$  can be neglected. For normal transistor operation, the voltage dependence of  $w_{BE,b}$  and the value of  $w_{BE,b}$  itself are usually fairly small compared to  $w_{Bm}$  and  $w_{BC,b}$  due to the forward biased junction and the fact that only fairly small  $V_{B'E'}$  changes are required to obtain large current variations. In contrast, variations in  $V_{B'C'}$  are often much larger. This and the tendency towards increasing collector doping for achieving high-speed performance cause the voltage dependence of  $w_{BC,b}$  to be the major factor in the bias variation of  $w_B$  under circuit relevant operation. This "dynamic" Early-effect can be described analytically as [38]

$$w_B \approx w_{B0} [1 - k_b(c - 1)] \quad (3.177)$$

with the BC capacitance ratio  $c = C_{jCi0}/C_{jCi}(V_{B'C'})$  and the width ratio  $k_b = w_{BC,b0}/w_{B0}$ ; here,  $w_{BC,b0}$  and  $w_{B0}$ , respectively, is the zero-bias width of the base sided BC SCR and neutral base region, respectively.

Besides  $w_B$  the second term in (3.169) contains the carrier velocity  $v_c$  in the BC SCR, which depends on the electric field in the BC junction. For not too low collector voltages  $V_{ci}$  the electric field is high enough to maintain saturation velocity. However, below a certain value of  $V_{ci}$  the drop of the velocity cannot be neglected anymore, leading to an increase in the storage time portion  $\tau_{Bfv}$  according to (3.174). The field dependence of the velocity in silicon can be roughly approximated by

$$v_c = v_{sn} \frac{u}{1 + u} \quad (3.178)$$

with the field ratio

$$u = \frac{E_{jc}(V_{B'C}, I_T)}{E_{lim}} \quad (3.179)$$

and  $E_{jc}$  as the field at the BC junction  $x = x_{jc}$  (cf. section 3.2). In the low-injection region considered here the current dependence of  $E_{jc}$  is negligible. Furthermore, the impact of the field in (3.178) is only relevant once it drops to a value around  $E_{lim}$  and below. Hence, for practical collector doping profiles the low-voltage solution (3.28) always applies which reads for the low-current case simply

$$E_{jc}(I_T = 0) = E_{jc0} = \sqrt{\frac{2qN_{Ci}}{\epsilon}} V_{ci}. \quad (3.180)$$

With  $V_{lim}$  from (3.9) one obtains after replacing the field variables

$$u = 2\sqrt{V_{PT}V_{ci}}/V_{lim}, \quad (3.181)$$

which yields a quite simple description for the voltage dependence of the carrier velocity  $v_c$  from (3.178).

The two variables  $w_B$  and  $v_c$  in (3.170) have a similar impact on the storage time for a given change of  $V_{B'C}$ . For instance, increasing  $V_{B'C}$  reduces the electric field in the BC SCR and its width. As a consequence, the width of the neutral base increases and the velocity decreases, both resulting in an increase of the storage time. Inserting (3.177) and (3.178) into (3.170) gives for the total bias dependent base storage time

$$\tau_{Bf} = \tau_{Bfd0}[1 - k_b(c - 1)]^2 + \frac{\tau_{Bfv0}[1 - k_b(c - 1)]}{u_0^{-1} + 1} \left(1 + \frac{1}{u}\right) \quad (3.182)$$

with the time constants at  $V_{B'C} = 0$  for the “diffusion” portion,

$$\tau_{Bfd0} = \frac{w_{B0}^2}{F_\zeta \bar{\mu}_{nB} V_T}, \quad (3.183)$$

and for the collector “exit” related portion

$$\tau_{Bfv0} = \frac{w_{B0}}{G_\zeta v_{sn}} \frac{1 + u_0}{u_0}. \quad (3.184)$$

### 3.5.1.2 Base-collector region

Although in classical theory for the analysis of the base region the carrier velocity at the end of the neutral base is assumed to be infinitely high, the importance of the transit time through the BC space-charge region had been recognized fairly early. For instance, [2] contains a derivation based on the impulse response of the depletion layer; the first-order approximation of the frequency dependent result gives the well-known relation

$$\tau_{BCv} = \frac{w_{BC}}{2v_s}, \quad (3.185)$$

where  $w_{BC}$  is the depletion layer width and  $v_s$  is the carrier saturation velocity. A more simple derivation for the general case of a partially depleted collector region and the relation to the small-signal variation of the total hole charge, which can be obtained from measurements, is given below.

First however, the electrical behavior of the base-collector region at low current densities shall be discussed qualitatively from a physics-based point of view. Figure 3.18 shows for both the case of a low and a high collector voltage the *small-signal* carrier distributions within the BC region for the BJT with the constant collector doping lowered to  $2 \cdot 10^{16} \text{cm}^{-3}$ . From the distribution for a small-signal  $V_{B'E'}$  short (lower figures) the ends of the respective BC SCR are clearly visible by the peaks. For the high-voltage case punch-through is reached, while for the low voltage case a significant portion of the collector is undepleted. The upper figures visualize the impact of a small-signal change in  $I_T$  at BC short. For the high voltage in (a) the mobile charge has the same impact (also on the field) as widening the SCR, which causes a reduction of electrons at its end. For the low voltage the mobile charge leads to a collapse of the SCR and, thus, an increase of electrons at its end. In both cases, the overall electron charge variation is compensated for by a hole charge increase in the base.

In the derivation below, the *small-signal* case is indicated by a “ $\delta$ ” in front of each variable. The derivation starts with the Poisson equation, which yields after integrating once from an arbitrary location  $x$  within the BC SCR to the end of the BC SCR for the electric field

$$\delta E(w_{BC}) - \delta E(x) = \frac{1}{\varepsilon} \int_x^{w_{BC}} \delta \rho(\xi) d\xi \quad (3.186)$$

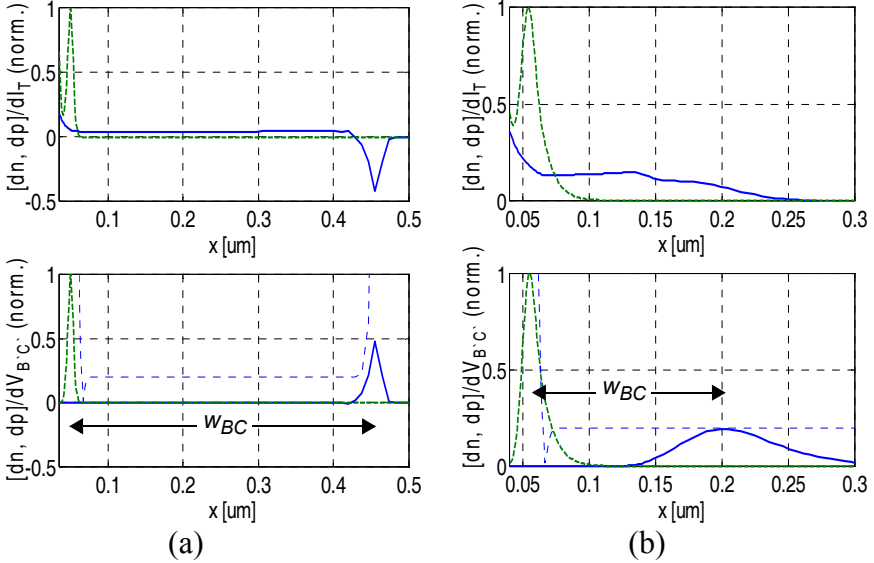


Fig. 3.18: Normalized small-signal distributions of electrons (solid line) and holes (dashed line) within the BC region of the BJT with  $N_{Ci} = 2 \cdot 10^{16} \text{ cm}^{-3}$ . (a) High-voltage (punch-through) case with  $(V_{B'E'}, V_{B'C'}) = (0.78, -3.3) \text{ V}$ . (b) Low-voltage case with  $(V_{B'E'}, V_{B'C'}) = (0.77, 0.4) \text{ V}$ . The upper figures show the changes w.r.t. to collector current at BC short, while the lower figures show the changes w.r.t. BC voltage at BE short. The dotted line represents the doping profile (not to scale).

with an arbitrary spatial dependence of the space charge  $\delta \rho$  and assuming a spatially independent permittivity. The width  $w_{BC}$  of the SCR corresponds to the value reached *after* the bias change. Generally, the small-signal width change is negligible compared to the change of the electrical variables (see e.g. [39, p. 31]). The field  $\delta E(w_{BC})$  is given by the voltage drop  $r_{Ci} \delta I_T$  across the non-depleted ohmic portion of the collector with resistivity  $\rho_{Ci}$  and associated resistance  $r_{Ci} = \rho_{Ci}(w_{Ci} - w_{BC})/A_E$ . Assuming a negligible extension of the SCR into the base and integrating again from the junction to the end of the BC SCR then gives the voltage across the SCR

$$\delta V = \delta V_{B'C} + r_{Ci} \delta I_T = - \int_0^{w_{BC}} \delta E(x) dx, \quad (3.187)$$

which also includes the voltage drop across the non-depleted ohmic portion. Inserting  $\delta E(x)$  from (3.186) leads to

$$\delta V = \frac{1}{\varepsilon} \int_0^{w_{BC}} \left( \int_x^{w_{BC}} \delta \rho(\xi) d\xi \right) dx - \int_0^{w_{BC}} \delta E(w_{BC}) dx \quad (3.188)$$

with  $\delta E(w_{BC}) = -\rho_{Ci} \delta I_T / A_E$ . The voltage

$$\int_0^{w_{BC}} \delta E(w_{BC}) dx = -\frac{\rho_{Ci} w_{BC}}{A_E} \delta I_T \quad (3.189)$$

corresponds to the field in the SCR caused by the voltage drop in the ohmic region. Combining the two  $\rho_{Ci}$  related terms on the l.h.s. and the r.h.s. of (3.188) gives

$$\varepsilon \delta V_c = \int_0^{w_{BC}} \left( \int_x^{w_{BC}} \delta \rho(\xi) d\xi \right) dx \quad (3.190)$$

with the modified voltage

$$\delta V_c = \delta V_{B'C} + r_{Ci0} \delta I_T, \quad (3.191)$$

and the low-field collector resistance

$$r_{Ci0} = \rho_{Ci} w_{Ci} / A_E. \quad (3.192)$$

Defining the functions  $f' = df/dx = 1$  and  $g = \int_x^{w_{BC}} \delta \rho(\xi) d\xi$ , and using  $(fg)' = fg' + f'g$  allows to write

$$\int_0^{w_{BC}} \left( \int_x^{w_{BC}} \delta \rho(\xi) d\xi \right) dx = \left( x \int_x^{w_{BC}} \delta \rho dx \right) \Big|_0^{w_{BC}} - \int_0^{w_{BC}} x \delta \rho dx.$$

The first expression on the r.h.s cancels after evaluation at the lower and upper bound. Extending the integrand of the remaining expression by  $w_{BC} \delta \rho$  gives:

$$\int_0^{w_{BC}} x \delta \rho(x) dx = w_{BC} \int_0^{w_{BC}} \delta \rho dx - w_{BC} \int_0^{w_{BC}} \left( 1 - \frac{x}{w_{BC}} \right) \delta \rho dx. \quad (3.193)$$

With the bias dependent charge stored within the collector-sided SCR,

$$\delta Q_{BC} = A_E \int_0^{w_{BC}} \delta \rho(x) dx, \quad (3.194)$$

inserting above expression with (3.193) into (3.190) gives after some rearrangement

$$A_E \frac{\varepsilon \delta V_c}{w_{BC}} = -\delta Q_{BC} + A_E \int_0^{w_{BC}} \left(1 - \frac{x}{w_{BC}}\right) \delta \rho dx. \quad (3.195)$$

Realizing that  $C_{jCi} = A_E(\varepsilon/w_{BC})$  is the BC depletion capacitance, the term on the l.h.s., after inserting (3.191) for  $\delta V_c$ , equals the sum of the small-signal BC depletion and current induced charge,

$$A_E \frac{\varepsilon \delta V_c}{w_{BC}} = \delta Q_{jCi} + r_{Ci0} C_{jCi} \delta I_T. \quad (3.196)$$

with  $\delta Q_{jCi} = C_{jCi} \delta V_{BC}$ . Inserting into (3.195) yields

$$-\delta Q_{BC} = \delta Q_{jCi} + r_{Ci0} C_{jCi} \delta I_T - A_E \int_0^{w_{BC}} \left(1 - \frac{x}{w_{BC}}\right) \delta \rho dx. \quad (3.197)$$

A similar result appears to have been derived in [39], but with two significant differences. First, the derivation assumes large-signal conditions forcing to define the depletion charge as  $C_{jCi}(V_{DCi} - V_{B'C'})$ , which is incorrect. Second, the  $r_{Ci0}$  related term resulting from the undepleted collector portion is missing.

Any hole charge variation  $dQ_{p,BC}$  on the base side caused by a change in  $V_{B'C'}$  or  $I_T$  (within  $\rho$ ) must be equal and opposite to the change  $dQ_{BC}$  (cf. Fig. 3.18). Thus, using (3.197) allows to write for the contribution of the collector region time constant to the transit frequency (defined in common-emitter configuration)

$$\left. \frac{dQ_{p,BC}}{dI_T} \right|_{V_{CE}} = C_{jCi} \left. \frac{dV_{BC}}{dI_T} \right|_{V_{CE}} + r_{Ci0} C_{jCi} + \tau_{BCv}. \quad (3.198)$$

with the time constant

$$\tau_{BCv} \cong -A_E \int_0^{w_{BC}} \left(1 - \frac{x}{w_{BC}}\right) \frac{d\delta \rho(x)}{dI_T} \Big|_{V_{CE}} dx. \quad (3.199)$$

In order to evaluate the integral term a specific expression for the space-charge has to be chosen. Assuming the most simple case of a constant doping profile and average carrier velocity  $v_c$  gives with  $I_T = qA_E n v_c$  the bias dependent expression

$$\rho(x) = q(N_{Ci} - n) = qN_{Ci} - I_T / (A_E v_c). \quad (3.200)$$

This results in the time constant

$$\tau_{BCv} = \frac{w_{BC}}{2v_c}, \quad (3.201)$$

which is the same result as (3.185) if  $v_c$  equals the saturation velocity. However, a spatially non-constant velocity leads to a different factor than 1/2.

Since  $dV_{B'C'} = dV_{B'E'}$  in common-emitter configuration, the derivative in (3.198)

$$\left. \frac{dV_{B'C'}}{dI_T} \right|_{V_{CE}} = \left. \frac{dV_{B'E'}}{dI_T} \right|_{V_{CE}} = g_{mi} \quad (3.202)$$

is equivalent to the transconductance of the intrinsic transistor. Inserting this into (3.198) yields

$$\left. \frac{dQ_{p,BC}}{dI_T} \right|_{V_{CE}} = C_{jCi} g_{mi} + r_{Ci0} C_{jCi} + \tau_{BCv}. \quad (3.203)$$

The first term,  $C_{jCi} g_{mi}$ , corresponds to the BC depletion charge variation, which is usually taken into account by compact model *equivalent circuits*. The last two terms can be lumped into the base-collector storage time

$$\tau_{BC} = r_{Ci0} C_{jCi} + \tau_{BCv} \quad (3.204)$$

and automatically included in the measured transit time.

Generally, the variables in  $\tau_{BC}$  are bias dependent and shall be discussed for the considered case of low current densities. A decreasing collector voltage  $V_{C'B'}$  reduces the electric field and the SCR width. As a consequence,  $C_{jCi}$  increases and so does the first term of (3.204). However, the decrease of both  $w_{BC}$  and  $v_c$  partially compensates each other in  $\tau_{BCv}$ . A compact analytical description of these bias dependences can be derived



on the basis of the existing theory but will be compact model specific since it requires smoothing functions are required to connect different regions of operation.

At this point it is instructional to take a look at the hole charge variations in common-base configuration. From (3.198) with  $\tau_{BCv}$  defined by (3.199) the desired small-signal expressions read

$$\left. \frac{dQ_{p,BC}}{dV_{B'E}} \right|_{V_{B'C}} = (r_{Ci0}C_{jCi} + \tau_{BCv})g_{mi,b} \quad (3.205)$$

$$\left. \frac{dQ_{p,BC}}{dV_{B'C}} \right|_{V_{B'E}} = C_{jCi} + (r_{Ci0}C_{jCi} + \tau_{BCv})g_{ci,b} \quad (3.206)$$

where

$$g_{mi,b} = \left. \frac{dI_T}{dV_{B'E}} \right|_{V_{B'C}} \quad \text{and} \quad g_{ci,b} = \left. \frac{dI_T}{dV_{B'C}} \right|_{V_{B'E}} \quad (3.207)$$

are the intrinsic transconductance and output conductance, respectively, in common-base configuration. Note that for most practical applications  $g_{mi,b} \cong g_{mi}$  and  $(r_{Ci0}C_{jCi} + \tau_{BCv})g_{ci,b} \ll C_{jCi}$ .

Above results have been obtained for a partially depleted collector region. In case of a fully depleted collector (punch-through) the voltage drop  $r_{Ci}\delta I_T$  is zero and, hence,  $\delta E(w_{Ci}) = 0$ . The time constant then reads

$$\tau_{BC} = \tau_{BCv}, \quad (3.208)$$

which equals (3.185) for sufficiently high electric fields with  $v_c = v_{sn}$ .

Portions of the results obtained in this section were derived in the literature. Eqs. (3.201) and (3.208) were also derived earlier in [40, 41] from simple quasi-static considerations. As a result of the quasi-static derivation, the time constants in  $\tau_{BC}$  can be calculated from device simulation results by simple and computationally inexpensive quasi-static regional analysis.

### 3.5.1.3 Neutral emitter

The stored charge that is relevant for high-frequency operation requires the calculation of the hole distribution within the emitter region. As discussed for the emitter back-injection current, the carrier distribution depends on the properties of the mono-silicon and poly-silicon region as well as on the interface between these two regions, the grain boundaries in the poly-silicon, and at the contact. In this section a practically suitable expression for the emitter storage time is derived based on the carrier distributions in Fig. 3.12 and the same assumptions as for the back-injection current in the mono-silicon region. However, for the charge also the contribution stored in the poly-silicon region needs to be taken into account (if existing). Therefore, the notion of an effective interface recombination velocity cannot be used here.

Generally, an analytical calculation of the hole density requires the solution of both the transport *and* the continuity equation of holes in a stack of mono- and poly-silicon regions representing the different grains, taking into account possible grain interface recombination and even tunneling at the poly-mono-silicon interface (e.g. [36]). Although such calculations provide insight into the dependence of the base current and charge on material and structural properties, the associated expressions are too complicated for compact models. Also, the involved material parameters are usually impossible to determine during a compact model parameter extraction. As a consequence, a more simple derivation is given below, that still resembles the basic effects and maintains an accuracy that is comparable with the uncertainties of the emitter material properties.

Neglecting volume recombination in the mono-silicon emitter due to the shallow junction depth gives for the excess hole distribution

$$\Delta p_m(x) = \frac{I_{jBEi} w_E}{q A_E \bar{\mu}_{pEm} V_T} \left( \frac{x}{w_E} \right) + \Delta p(0), \quad (3.209)$$

in which the prefactor of the  $x$ -dependent term results from the transport equation. In contrast, in the poly-silicon region width  $w_p$  is usually larger than the diffusion length  $L_{pp} = \sqrt{\bar{\mu}_{pEp} \tau_{pEp}}$ , with  $\bar{\mu}_{pEp}$  and  $\tau_{pEp}$  as the average velocity and lifetime of holes in poly-silicon. Thus, the solution of the second-order differential equation

$$\bar{\mu}_{pEp} V_T \frac{d^2 \Delta p}{dx^2} = \frac{\Delta p}{\tau_{pEp}} \quad (3.210)$$

is required, which results from inserting the transport equation along with all assumptions made above into the continuity equation. Considering the unknown process dependent properties of the poly-silicon it suffices to perform the solution for just a single region in order to obtain a rough approximation for the dependence of the total emitter charge on the charge stored in the poly-silicon region. Solving (3.210) with the boundary condition of negligible carrier density at the emitter metal contact (cf. Fig. 3.12) leads to the following result for the excess hole density

$$\Delta p_p(x) = \Delta p(0) \frac{\sinh([x + w_p]/L_{pp})}{\sinh(w_p/L_{pp})}. \quad (3.211)$$

Inserting this into the (diffusion) hole transport equation leads to

$$-J_{px}(0)A_E = -qA_E \bar{\mu}_{pEp} V_T \left. \frac{d\Delta p_p(x)}{dx} \right|_0 = qA_E \frac{\bar{\mu}_{pEp} V_T}{L_{pp}} \Delta p(0) \coth\left(\frac{w_p}{L_{pp}}\right). \quad (3.212)$$

The current on the l.h.s. is given by the interface recombination condition

$$-J_{px}(0)A_E = I_{jBEi} - qA_E v_{pmI} \Delta p(0), \quad (3.213)$$

with  $v_{pmI}$  as the poly-mono interface recombination velocity. Inserting into (3.212) and solving for  $\Delta p(0)$  yields

$$\Delta p(0) = \frac{1}{qA_E \frac{\bar{\mu}_{pEp} V_T}{L_{pp}} \coth\left(\frac{w_p}{L_{pp}}\right) + v_{pmI}} I_{jBEi}. \quad (3.214)$$

Inserting this into (3.209) and (3.211) gives the hole distribution as a function of the injected base current.

Integration of (3.209) leads to the stored charge in the mono-silicon emitter region

$$Q_{pEm} = qA_E \int_0^{w_E} \Delta p_m(x) dx = \frac{w_E^2 I_{jBEi}}{2 \bar{\mu}_{pEm} V_T} + qA_E \Delta p(0) w_E \quad (3.215)$$

and for the stored charge in the poly-silicon region

$$Q_{pEp} = qA_E \int_{-w_p}^0 \Delta p dx = qA_E \Delta p(0) L_{pp} \frac{\cosh\left(\frac{w_p}{L_{pp}}\right) - 1}{\sinh\left(\frac{w_p}{L_{pp}}\right)}. \quad (3.216)$$

The total stored emitter charge  $Q_{pE}$  is the sum of the above.

The emitter storage time in compact transport models is defined as

$$\tau_{Ef} = \left. \frac{dQ_{pE}}{dI_T} \right|_{V_{CE}} = \left. \frac{dQ_{pE}}{dI_{jBEi}} \right|_{V_{CE}} \left. \frac{dI_{jBEi}}{dI_T} \right|_{V_{CE}} = \left. \frac{dQ_{pE}}{dI_{jBEi}} \right|_{V_{CE}} \frac{1}{\beta_{f0}} \quad (3.217)$$

with  $\beta_{f0}$  as the low-frequency small-signal forward current gain in common-emitter configuration. Obviously, the storage time consists of two parts. The first one is associated with the mono-silicon region:

$$\tau_{Efm} = \frac{1}{\beta_{f0}} \left( \frac{w_E^2}{2\bar{\mu}_{pEm} V_T} + \frac{w_E}{\frac{\bar{\mu}_{pEp} V_T}{L_{pp}} \coth\left(\frac{w_p}{L_{pp}}\right) + v_{pmI}} \right). \quad (3.218)$$

Letting  $w_p/L_{pp} \rightarrow 0$  would imply an infinite gradient of  $\Delta p(x)$  unless  $\Delta p(0)$  is reduced to accommodate the finite value of  $I_{jBEi}$  at a given velocity and vice versa. The extreme case of infinite  $v_{pmI}$  leads to  $\Delta p(0) = 0$ . Therefore,  $w_p/L_{pp} \rightarrow 0$  causes the poly-silicon related term to disappear rather than to approach  $w_E/v_{pmI}$ . The interface and its associated condition (3.213) now needs to be replaced by a contact with the condition  $I_{p,con} = I_{jBEi} = qA_E v_{Epm} \Delta p(0)$  according to (3.98).

Using (3.216), the contribution of the poly-silicon region reads

$$\tau_{Efp} = \frac{1}{\beta_{f0}} \left( \frac{L_{pp}}{\frac{\bar{\mu}_{pEp} V_T}{L_{pp}} \coth\left(\frac{w_p}{L_{pp}}\right) + v_{pmI}} \frac{\cosh(w_p/L_{pp}) - 1}{\sinh(w_p/L_{pp})} \right). \quad (3.219)$$

Series expansion of the hyperbolic functions for the case of  $w_p/L_{pp} \rightarrow 0$  correctly yields  $\tau_{Efp} \rightarrow 0$ .

In summary, the emitter storage time for a two-region approach reads

$$\tau_{Ef} = \tau_{Efm} + \tau_{Efp}. \quad (3.220)$$

For a sufficiently large value of  $v_{pmI}$  the expression simplifies to

$$\tau_{Ef} = \frac{1}{\beta_{f0}} \left( \frac{w_E^2}{2\bar{\mu}_{pEm} V_T} + \frac{w_E + L_{pp} \frac{\cosh(w_p/L_{pp}) - 1}{\sinh(w_p/L_{pp})}}{v_{pmI}} \right). \quad (3.221)$$

Very often, a more simple expression is found which is derived under the assumptions made for the base current derivation; i.e. with the effects at the interface and in the poly-silicon region represented by an effective recombination velocity. Generally, in the derivation the boundary condition (3.98) with the current density related recombination velocity  $v_{Epm}$  is used leading to

$$\tau_{Ef} = \frac{1}{\beta_{f0}} \left[ \frac{w_E^2}{2\bar{\mu}_{pEm} V_T} + \frac{w_E}{v_{Epm}} \right], \quad (3.222)$$

This simple representation can be maintained if a charge related *effective* recombination velocity is defined from (3.221) as

$$v_{Epm} = \frac{v_{pmI}}{1 + \frac{L_{pp}}{w_E} \frac{\cosh(w_p/L_{pp}) - 1}{\sinh(w_p/L_{pp})}}. \quad (3.223)$$

The assumption of a spatially independent hole current density in the mono-silicon emitter is (even in the first SiGe-HBT generations) not always justified and introduces a certain amount of error in the theoretical expression. Also note that in HBTs carrier injection into the emitter is strongly suppressed up to very high current densities. As a consequence, emitter charge storage becomes negligible, reducing the importance of the corresponding modeling.

### 3.5.1.4 Base-emitter space charge region

Classical theory assumes that the SCR does not contain any mobile carriers. In reality though the *minority* carrier concentration is largest around the junction, as is sketched Fig. 3.19. Based on the assumption of a very small BE SCR width, this so-called neutral charge [42] is usually neglected. However, according to Fig. 2.12 the storage time  $\tau_{BE}$  of the BE SCR is the dominant contribution to  $\tau_{mS}$  at (very) low current densities where it causes the increase of  $\tau_f$ . It also interferes with the determination of the BE depletion capacitance from transit frequency measurements [43]. Below an analytical approximation is derived by means of Fig. 3.19. For simplifying the derivation, which was first given in [44] the origin of the  $x$ -axis is moved to  $x_{je}$ .

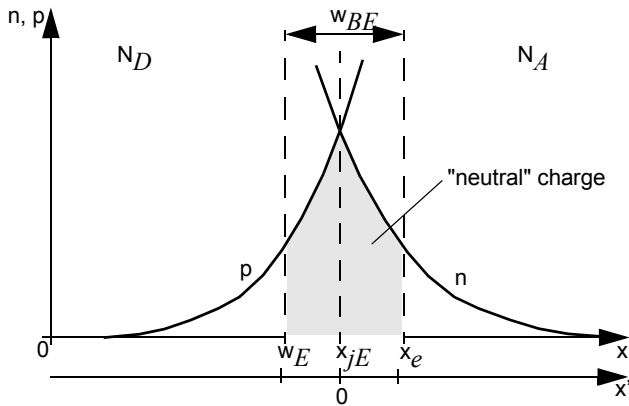


Fig. 3.19: Sketch of carrier densities in the BE SCR used for deriving an analytical expression for the neutral charge and related storage time.

The first step is to calculate the current dependence of the neutral charge. For forward operation mode, which is of interest here, the SCR width  $w_{BE}$  is fairly small so that an average value  $\bar{n}_i$  for the intrinsic carrier concentration and spatially constant values for the quasi-Fermi potentials are assumed within the SCR. Then, the  $pn$  product reads

$$p(x)n(x) = \bar{n}_i^2 \exp\left(\frac{V_{BE}}{V_T}\right). \quad (3.224)$$

Around the junction, i.e. within  $w_{BE}$ , it is reasonable to assume a symmetric spatial dependence of  $n$ ,  $p$  and doping  $N$ . Hence,  $N(x) = a x'$  is a good approximation. As a consequence of symmetry and (3.224), the carrier densities at  $x' = 0$  read

$$p(0) = n(0) = \bar{n}_i \exp\left(\frac{V_{B'E'}}{2V_T}\right), \quad (3.225)$$

while the spatial dependence of the carrier densities is to the first order exponential,

$$p(x') = p(0) \exp\left(\frac{x'}{x_L}\right), \quad n(x) = n(0) \exp\left(-\frac{x'}{x_L}\right), \quad (3.226)$$

with  $x_L$  as a characteristic length that will be determined next.

The minority carrier densities at the SCR edges are given by the neutrality condition,

$$p\left(-\frac{w_{BE}}{2}\right) = \frac{\bar{n}_i^2}{a(w_{BE}/2)} \exp\left(\frac{V_{B'E'}}{V_T}\right) = n\left(\frac{w_{BE}}{2}\right). \quad (3.227)$$

In what follows, it suffices to consider only the right half of the SCR due to the symmetry. Inserting  $w_{BE}/2$  into (3.226) yields with (3.227) and (3.225) the characteristic length

$$x_L = \frac{w_{BE}/2}{\ln\left(a \frac{w_{BE}}{2\bar{n}_i}\right) - \frac{V_{B'E'}}{2V_T}}. \quad (3.228)$$

Recognizing that for a linear junction (e.g. [25])

$$V_{DEi} = 2V_T \ln\left(a \frac{w_{BE}}{2\bar{n}_i}\right), \quad (3.229)$$

is the built-in voltage and inserting  $x_L$  back into (3.226) gives for the spatial dependence of the carrier density

$$n(x') = \bar{n}_i \exp\left(\frac{V_{B'E'}}{2V_T}\right) \exp\left(-\frac{x'}{w_{BE}} \frac{V_{DEi} - V_{B'E'}}{V_T}\right). \quad (3.230)$$

The total neutral charge  $Q_{BE}$  is obtained by integrating the carrier density over  $x'$  and multiplying by 2 due to symmetry:

$$Q_{BE} = 2qA_E V_T \frac{w_{BE}}{V_{DEi} - V_{B'E'}} \bar{n}_i \left[ 1 - \exp\left(-\frac{V_{DEi} - V_{B'E'}}{V_T}\right) \right] \exp\left(\frac{V_{B'E'}}{2V_T}\right). \quad (3.231)$$

In the bias range, where  $\tau_{BE}$  is relevant, the  $\exp()$  term in the brackets is much smaller than 1, so that after inserting the voltage dependence  $w_{BE} = w_{BE0}(1 - V_{B'E'}/V_{DEi})^{-z_{Ei}}$  the result reduces to

$$Q_{BE} = \frac{Q_{BE0}}{(1 - V_{B'E'}/V_{DEi})^{(1+z_{Ei})}} \exp\left(\frac{V_{B'E'}}{2V_T}\right). \quad (3.232)$$

with the charge parameter

$$Q_{BE0} = 2qA_E \bar{n}_i w_{BE0} (V_T/V_{DEi}). \quad (3.233)$$

At low injection, (3.232) gives roughly a dependence with  $I_T^{1/2}$ .

The corresponding storage time is given by the derivative of  $Q_{BE}$  w.r.t. the transfer current:

$$\tau_{BE} \cong \left. \frac{dQ_{BE}}{dV_{B'E'}} \frac{dV_{B'E'}}{dI_T} \right|_{V_{CE}} = \frac{Q_{BE}}{2V_T} \left( 1 + \frac{2V_T}{V_{DEi}} \frac{1+z_{Ei}}{1 - \frac{V_{B'E'}}{V_{DEi}}} \right) \frac{1}{g_{mi}}. \quad (3.234)$$

From above expression it becomes clear that  $\tau_{BE}$  decreases rapidly with increasing collector current since  $g_{mi} \sim I_C$ . Note that the term  $Q_{BE0}/(2V_T)$  has the dimension of a capacitance with an area specific value in the order of several fF/ $\mu\text{m}^2$  depending on the process technology.

In order to assess the accuracy of the derived formulation, Fig. 3.20 shows a comparison between the analytical expression (3.234) and the results obtained from a regional analysis of the 1D BJT. Excellent agreement is obtained after determining the parameter  $Q_{BE0}$  at a bias point at low currents, e.g.,  $V_{B'E'} = 0.7\text{V}$ , the choice of which is not sensitive in contrast to the value of  $V_{DEi}$ . The strong increase at high currents is caused by the drop of the transconductance there. A simple approximation can be obtained by lumping the bias dependent terms in (3.232) and (3.234) that re-



late to  $V_{DEi}$  into a modified exponent coefficient  $m_{tBE} < 2$ , leading to the charge

$$Q_{BE} \cong Q_{BE0} \exp\left(\frac{V_{B'E'}}{m_{\tau BE} V_T}\right) \quad (3.235)$$

and the simplified expression for the storage time

$$\tau_{BE} \cong \frac{Q_{BE}}{m_{\tau BE} V_T g_{mi}}. \quad (3.236)$$

As the dashed line in Fig. 3.20 shows, above expression still gives reasonable agreement if  $Q_{BE0}$  is determined at a higher bias point than for (3.234).

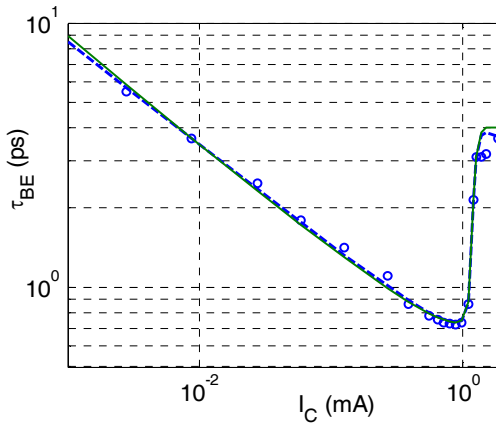


Fig. 3.20: Neutral BE storage time vs. collector current: comparison between eq. (3.234) (solid line), simplified expression (3.236) (dashed line) and results from 1D BJT device simulation using the Regional Approach (symbols). Model parameters:  $V_{DEi} = 1.2\text{V}$ ,  $z_{Ei} = 0.333$ ,  $Q_{BE0} = 4.7\text{fC}$ ;  $m_{\tau be} = 1.6$ ;  $V_{B'C'} = 0\text{V}$ .

The main problem though is to determine  $\tau_{BE}$  from measurements since at low current densities its bias variation is masked by the much larger term  $C_B/g_{mi}$ , with  $C_B$  as the sum of all (depletion and parasitic) capacitances connected to the base terminal (cf. sec. 2.4.2). Therefore, in compact models an explicit description of  $Q_{BE}$  and  $\tau_{BE}$  does usually not exist and the contribution is absorbed by the existing minority and depletion charge related terms.

### 3.5.1.5 Summary of the results for low current densities

The sum

$$\tau_{m\Sigma} = \tau_{Ef} + \tau_{BE} + \tau_{Bf} + \tau_{BC} \quad (3.237)$$

is a physics-based approximation of the measured transit time  $\tau_f$ . Except for  $\tau_{BE}$  all other components depend mainly on the collector voltage through the width variation of the neutral base (via  $\beta_f$  also in  $\tau_{Ef}$ ) and the BC SCR as well as through the carrier velocity in the BC SCR. The derived equations results from a derivative of the total mobile charge at forward operation w.r.t. the transfer current. This includes the implicit assumption that the forward transfer current  $I_{Tf}$  equals the total transfer current, which is the case as long as  $\exp(V_{C'E}/V_T) \gg 1$ . At hard saturation, when this condition is not fulfilled anymore,  $I_{Tf}$  has to be used in the equations rather than  $I_T$ . In addition, the derivative of the total mobile charge w.r.t. the inverse (or reverse) transfer current component  $I_{Tr}$  becomes relevant, which then increases rapidly. This derivative corresponds to an inverse (or reverse) time constant  $\tau_r$  which represents the strong dependence of the charge w.r.t.  $V_{B'C'}$  under the conditions of low and negative  $V_{C'E'}$ .

### 3.5.2 Medium current densities

As the previous discussion has shown, the subdivision into different current density ranges is determined by the mechanisms occurring in the collector region. The most important variable is the electric field at (or around) the BC junction. Its current dependence was shown in Fig. 3.8 for different collector voltages.

Once the field starts to drop the storage times start to increase. Since there is no abrupt change in the field and velocity with increasing current there is also no exact definition of the boundary between low and medium current densities. As a consequence, the charge storage expressions derived for the low current region remain the basis for discussing and describing the medium current range. On the other end, the boundary between medium and high current densities can be defined more clearly based on the electric field and carrier distribution in the collector. The corresponding critical current will be derived in this section.

### 3.5.2.1 Neutral base region

The drop of the electric field  $E_{jc}$  causes  $w_B$  and  $v_n$  to change with current according to (3.176)-(3.179). Inserting the current dependent formulations for  $E_{jc}$  obtained in section 3.2 into the field dependent variables ( $w_B$ ,  $\zeta$ ,  $v_n$ ) in the base charge (3.169) would in principle be a solution, assuming the existence of a continuously differentiable  $E_{jc}$  description (cf. [12]). However, the resulting charge relation as well as the associated storage time would be far too complicated for a compact model. At the considered medium current densities it is reasonable to assume that the impact of a  $w_B$  change on  $\zeta$  and  $f_\zeta$  is negligible. This assumption, which is valid in particular for drift-type HBTs, simplifies the analytical expressions significantly since the factors  $F_\zeta$  and  $G_\zeta$  from (3.172) and (3.175), respectively, can be considered as constants. Indicating the remaining current dependent variables by writing down explicitly their argument ( $I_T$ ), the base charge can be written as

$$Q_{nB} = \left[ \frac{w_B^2(I_T)}{F_\zeta \bar{\mu}_{nB} V_T} + \frac{w_B(I_T)}{G_\zeta v_c(I_T)} \right] I_T. \quad (3.238)$$

The derivative still yields a drift-diffusion and a carrier jam related storage time. The first one now reads

$$\tau_{Bfd} = \frac{w_B^2}{F_\zeta \bar{\mu}_{nB} V_T} \left( 1 + 2 \frac{I_T}{w_B} \frac{dw_B}{dI_T} \right) \quad (3.239)$$

while one obtains for the second one

$$\tau_{Bfv} = \frac{w_B}{G_\zeta v_{cn}} \left( 1 + \frac{I_T}{w_B} \frac{dw_B}{dI_T} - \frac{I_T}{v_c} \frac{dv_c}{dI_T} \right). \quad (3.240)$$

Hence, besides a bias dependence in the low-current terms via  $w_B$  and  $v_{cn}$ , additional terms containing the derivative of these variables come into play. Since in both cases the electric field in the collector, or better its reduction with current, causes the additional terms, they can be expressed by the derivative of the field using the chain rule, yielding

$$\tau_{Bfd} = \frac{w_B^2}{F_\zeta \bar{\mu}_{nB} V_T} \left[ 1 + 2 \frac{I_T}{w_B} \frac{dw_B}{dE_{jc}} \frac{dE_{jc}}{dI_T} \right]_{V_{CE}} \quad (3.241)$$

and

$$\tau_{Bfv} = \frac{w_B}{G_\zeta v_c} \left[ 1 + I_T \left( \frac{1}{w_B} \frac{dw_B}{dE_{jc}} - \frac{1}{v_c} \frac{dv_c}{dE_{jc}} \right) \frac{dE_{jc}}{dI_T} \right]_{V_{CE}}. \quad (3.242)$$

Note that  $dw_B/dE_{jc} < 0$  so that with  $dv_c/dE_{jc} > 0$  and  $dE_{jc}/dI_T < 0$  the second term in the brackets is larger than zero. The detailed form of the derivative terms depends on the analytical description used for  $w_B$  and  $v_c$ . Generally, the current dependence of  $\tau_{Bfd}$  is relatively small compared to that of  $\tau_{Bfv}$ , since the base width change is fairly limited - especially for advanced BJTs and HBTs with their high base doping concentration.

### 3.5.2.2 Collector region

Assuming a one-sided abrupt BC junction and an ideal  $v_c(E)$  relationship, the width of the BC SCR can be calculated as a function of operating point from (3.27), the numerator  $V_{ci} - E_{wc} w_{Ci}$  of which is also current dependent via (3.20). Defining the *low-field* internal collector resistance

$$r_{Ci0} = \frac{1}{q N_{Ci} \mu_{nC}(N_{Ci}, E = 0)} \frac{w_{Ci}}{A_E} = \frac{V_{lim}}{I_{lim}} \quad (3.243)$$

and the "ohmic" critical current

$$I_{CKo} = V_{ci}/r_{Ci0}, \quad (3.244)$$

one obtains for the width after replacing the numerator in (3.27) and introducing the punch-through voltage

$$w_{BC,c} = w_{Ci} \sqrt{\frac{1 - \frac{I_T}{I_{CKo}}}{1 - \frac{I_T}{I_{lim}}}} \frac{V_{ci}}{V_{PT}} \frac{I_{CKo} \sqrt{1 - (I_T/I_{lim})^2}}{1 - I_T/I_{lim}}. \quad (3.245)$$

If the term  $\sqrt{1 - (I_T/I_{lim})^2}$  is neglected and  $w_{BC,c}(I_T = 0)$  is inserted, the result reported first in [10] is found

$$w_{BC,c} = w_{BC,c}(I_T=0) \sqrt{\frac{1 - I_T/I_{CKO}}{1 - I_T/I_{lim}}}. \quad (3.246)$$

Notice that  $I_{CKO} < I_{lim}$  for low voltages, while  $I_{CKO} > I_{lim}$  for high voltages. Therefore, according to the equations above the BC SCR collapses for low voltages while it expands toward the buried layer for sufficiently large voltages. This causes  $\tau_{BC}$  to become not only voltage but also current dependent.

Although (3.245) is an important result for obtaining quantitative insight into the current dependence of the BC SCR, it is obviously not suitable for compact models in its existing form since either the numerator or the denominator becomes zero above a certain current. Avoiding a negative square root and division by zero as well as limiting  $w_{BC,c}$  to  $w_{Ci}$  requires more or less sophisticated smoothing functions. If  $w_{BC,c}$  is also used for modeling the current dependence of the depletion capacitance an integration is necessary to obtain the charge; such an integration will be difficult due to the smoothing functions. In summary, for this medium current density range there are no suitable analytical charge expressions and associated accurate derivatives (capacitances, storage times) known for the neutral base and BC SCR region, that are based on classical transistor theory or can be used as extension thereof. Using a different approach though, it is possible to derive analytical equations for the important storage times, but the related discussion is deferred to sec. 4.4.

At this point, analytical expressions are derived only for the critical current that defines the boundary between medium- and high-current region.

### A Low collector voltage

As mentioned in the qualitative discussion earlier, the bias point at which the electric field in the collector becomes horizontal defines the critical current which is then given by (3.12). The field dependent mobility occurring in (3.12) can be written as

$$\mu_{nC_i}(N_{C_i}, E_{CKI}) = \frac{\mu_{nC_i0}}{[1 + (E_{CKI}/E_{lim})^{\alpha_\mu}]^{1/\alpha_\mu}} \quad (3.247)$$

with  $\mu_{nCi0}$  as the low-field mobility and  $\alpha_m (= 1 \dots 2)$  as coefficient. With the critical voltage  $V_{lim}$  from (3.9) and inserting  $E_{CKI}$  from (3.11) the mobility can be expressed as a function of the internal terminal voltage:

$$\mu_{nCi} = \frac{\mu_{nCi0}}{[1 + (V_{ci}/V_{lim})^{\alpha_\mu}]^{1/\alpha_\mu}}. \quad (3.248)$$

Inserting this as well as the low-field internal collector resistance defined by (3.243) and  $E_{CKI}$  into (3.12) yields for the *low-voltage critical current*

$$I_{CKI} = \frac{V_{ci}}{r_{Ci0}} \frac{1}{[1 + (V_{ci}/V_{lim})^{\alpha_\mu}]^{1/\alpha_\mu}}. \quad (3.249)$$

## B High collector voltage

The critical current density is defined for the situation when the electric field at the BC junction reaches  $E_{lim}$ , which separates ohmic and saturation region in the velocity-field curve. Assuming a spatially constant doping in the collector, integrating Poisson's equation twice and solving for the case  $E(0) = -E_{lim}$  at  $I_T = I_{CK}$  yields (3.8) for the *high-voltage critical current*, which is repeated here for convenience

$$I_{CKh} = I_{lim} \left[ 1 + \frac{V_{ci} - V_{lim}}{V_{PT}} \right]. \quad (3.250)$$

## C Critical current

The different components of the critical current derived above are shown in Fig. 3.21, along with the relevant parameters and their meaning in shaping the voltage dependence. The low-current component follows directly from the velocity-field curve in which the field axis and velocity axis, respectively, have been converted to a voltage axis and a current axis, respectively. The slope at very low voltages is determined by the low-field internal collector resistance. The high-voltage component simply represents the linear dependence of the electric field with applied voltage at punch through condition, with the slope being determined by the punch-through voltage.

In classical theory [7, 8] the critical electric field is set to zero, which corresponds to  $V_{lim} = 0$ . As can be seen in curve 4 in Fig. 3.2,  $v_c$  has strongly decreased already at electric field values below  $V_{lim}$ , which are still larger than zero. Also, the electron density possesses already a significant gradient, which indicates a significant diffusion current component. As a consequence, the above assumption results in values for the critical current density that are way too high to be useful for a proper indication of the onset of collector charge storage (i.e. high-current effects). Besides that, even at highest current densities  $E_{jc}$  approaches the finite value  $2V_T/w_{Ci}$  (cf. sec. 4.4).

As Fig. 3.21 shows the two components assume different values at  $V_{lim}$ . For use in a compact model this discontinuity has to be eliminated, e.g., by means of a suitable smoothing function. It should be mentioned that in high-speed transistors  $V_{lim}$  is typically of the order of 0.4...0.6V so that in practice most transistors are operated in the “high-voltage” range.

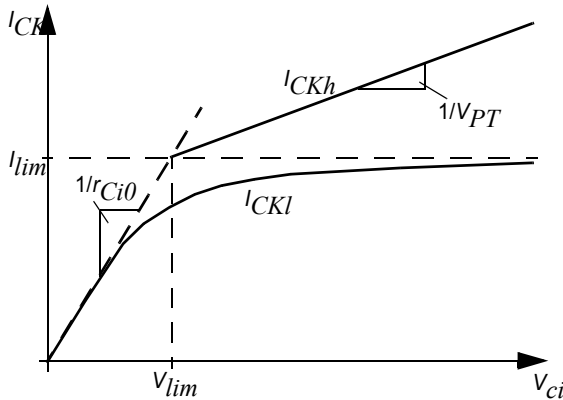


Fig. 3.21: Sketch of the critical current components as a function of collector voltage with associated parameters.

### 3.5.2.3 BC space-charge region

The corresponding storage time  $\tau_{BC}$  depends on both the BC SCR width  $w_{BC}$  and the carrier velocity  $v_c$ . According to (3.245) and assuming a partially depleted collector at  $I_T = 0$ ,  $w_{BC}$  can decrease or increase with current depending on the applied voltage, and so does  $\tau_{BC}$ . In addition, once

the electric field drops to values in the order of  $E_{lim}$  the velocity will decrease noticeably. This always leads to an increase of  $\tau_{BC}$ , which thus enhances its increase at larger collector voltages but compensates partially for the decrease at low voltages.

An analytical expression can be found by inserting (3.200) into (3.199),

$$\tau_{BCv} = \int_0^{w_{BC}} \left(1 - \frac{x}{w_{BC}}\right) \frac{d}{dI_T} \left(\frac{I_T}{v_c(x)}\right) \bigg|_{V_{CE}} dx, \quad (3.251)$$

which now requires to include the derivative of the current dependent velocity and the spatial dependence of the latter. A compact analytical description obviously requires approximations of  $w_{BC}$  and  $v_c$  as a function of voltage *and* current. This is possible with the equations derived so far, but only for certain operating regions. *Continuously differentiable* expressions require both simplifications and smoothing between the operating regions, which depend on the particular compact model formulation.

### 3.5.2.4 Neutral emitter region

The reason for the slight increase of  $\tau_{Ef}$  is the current dependence of the small-signal low-frequency current gain  $\beta_{f0}$ . Expressing the latter by the respective conductances and normalizing to a reference value at low current densities (index "l") and  $V_{B'C'} = 0$  gives

$$\frac{\beta_{f0}}{\beta_{f0,l}} = \frac{g_{mi}/g_{bei}}{g_{mi,l}/g_{bei,l}} = \frac{g_{mi}}{g_{mi,l}} \exp\left(\frac{V_{B'E,l} - V_{B'E'}}{m_{BEi} V_T}\right), \quad (3.252)$$

In BJTs the factor  $m_{BEi}$  is very close to 1 up to high current densities (cf. Fig. 3.13) and is not affected by the effects in the collector junction. Notice that the reference point should not be at too low current densities where the current gain drops again due to the influence of recombination. Any change of  $\beta_{f0}$  in that region is irrelevant though for modeling  $\tau_{Ef}$  since there (i) the emitter minority charge contribution is negligible and (ii)  $\tau_{Ef} \ll \tau_{BE}$  holds.

According to the GICCR (see sec. 4.3) the intrinsic transconductance in the bias range considered here can be written as



$$g_{mi} \cong \frac{I_{Tf} Q_{p,T} - (h_{jEi} C_{jEi} + h_{jCi} C_{jCi}) V_T}{V_T Q_{p,T} + \tau_{fT} I_{Tf}}, \quad (3.253)$$

with  $I_{Tf} = I_T \gg I_{Tr}$  for practical applications,  $Q_{p,T}$  and  $\tau_{fT}$  as transfer current related hole charge and (forward) storage time, respectively, and  $h_j$  as HBT related weighting factors ( $= 1$  for BJTs). The emitter storage time is proportional to the reciprocal value of  $\beta_{f0}$ , so that with  $I_{Tf} = (const/Q_{p,T}) \exp(V_{BE}/V_T)$  and  $m_{BEi} = 1$ , the ratio

$$\frac{\beta_{f0,l}}{\beta_{f0}} \cong \frac{Q_{p,T,l} + Q_{f,T}}{Q_{p,T,l}} \frac{Q_{p,T,l} + Q_{f,T} + \tau_{fT} I_{Tf}}{Q_{p,T,l}} \frac{Q_{p,T,l} - (h_{jEi} C_{jEi})|_l V_T}{Q_{p,T,l} + Q_{f,T} - h_{jEi} C_{jEi} V_T} \quad (3.254)$$

needs to be considered. Here  $Q_{f,T}$  is transfer current related weighted charge and  $\tau_{fT} I_{Tf}|_l \ll Q_{p,T,l}$ . The term  $h_{jEi} C_{jEi}$  does not change much over current and is also small compared to  $Q_{p,T,l}$  so that the first and last ratio of the r.h.s. cancel, leading to

$$\frac{\beta_{f0,l}}{\beta_{f0}} \cong \frac{Q_{p,T,l} + Q_{f,T} + \tau_{fT} I_{Tf}}{Q_{p,T,l}} = 1 + \frac{Q_{f,T} + \tau_{fT} I_{Tf}}{Q_{p,T,l}}. \quad (3.255)$$

A rough approximation of the r.h.s. for practical applications is

$$1 + \frac{Q_{f,T} + \tau_{fT} I_{Tf}}{Q_{p,T,l}} \approx 1 + \left( \frac{I_{Tf}}{I_{CK}} \right)^{g_{tE}} \quad (3.256)$$

where  $g_{tE}$  is a constant. Figure 3.22a shows a comparison of this approximation with the actual current gain ratio  $\beta_{f0,l}/\beta_{f0}$  using technological values for calculating the parameters of the critical current. As expected, the behavior is roughly captured. Setting  $g_{tE} = 2$  yields reasonable agreement for the slow increase at lower current densities while  $g_{tE} = 4$  turns out to provide a better fit at higher current densities.

Defining the bias independent low-current emitter storage time according to (3.222) by

$$\tau_{E f0} = \frac{1}{\beta_{f0,l}} \left[ \frac{w_E^2}{2 \bar{\mu}_{pE} V_T} + \frac{w_E}{v_{Epm}} \right] \quad (3.257)$$

yields after inserting (3.256) and (3.255) for the bias dependent emitter storage time [38]

$$\tau_{Ef} = \tau_{Ef0} \left[ 1 + \left( \frac{I_{Tf}}{I_{CK}} \right)^{g_{\tau E}} \right]. \quad (3.258)$$

Figure 3.22b shows a comparison of this expression with the actual emitter storage time from device simulation using the Regional Approach. The behavior before the sharp increase is reasonably well captured using  $g_{\tau E} = 4$ , but the agreement at high current densities is less satisfying in contrast to results for older technologies (e.g. [44, 38]). On the other hand, the importance of emitter charge storage is diminishing, especially for HBTs so that the value of a more elaborate analytical model is questionable.

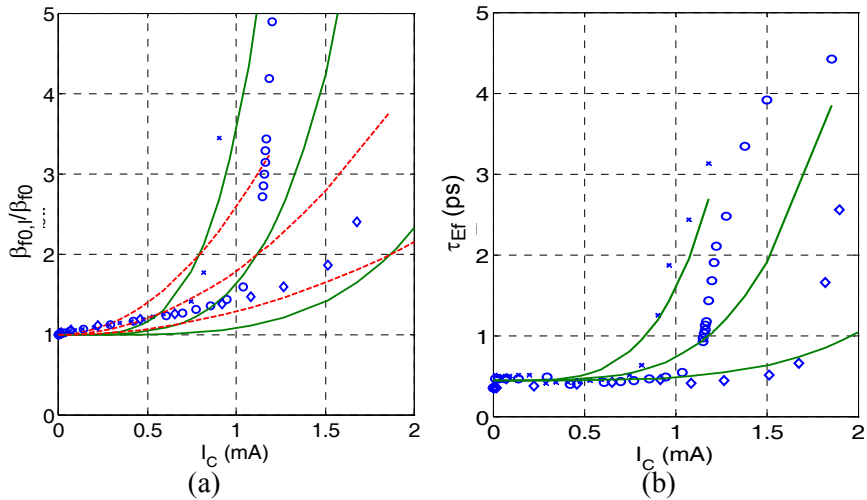


Fig. 3.22: Device simulation results for the emitter storage time of the 1D BJT for  $V_{B'C'}/V = 0.4, 0, -2.3$ : (a) Reciprocal of current gain ratio from device simulation (symbols) compared to expression (3.256) with  $g_{\tau E} = 4$  (solid lines) and  $g_{\tau E} = 2$  (dashed lines). (b) Comparison of the emitter storage time calculated from (3.258) with the Regional Approach result and  $g_{\tau E} = 4$ .

The approximation (3.258) allows to obtain an analytical expression for the corresponding charge in the neutral emitter region:

$$Q_{Ef} = \int_0^{I_{Tf}} \tau_{Ef} dI = \tau_{Ef0} I_{Tf} \left[ 1 + \frac{(I_{Tf}/I_{CK})^{g_{\tau E}}}{1 + g_{\tau E}} \right]. \quad (3.259)$$

### 3.5.2.5 BE space-charge region

The onset of the field reduction at the BC junction and associated increase in the total hole charge leads to a slower increase of  $I_T$  with  $V_{B'E'}$ , and, hence also reduction of the transconductance. According to (3.234), this leads to a weaker decrease of the BE SCR storage time than with  $1/\sqrt{I_T}$  (cf. Fig. 3.20).

### 3.5.2.6 Summary

With a proper analytical approximation of the current dependence of  $w_B$ ,  $v_c$  and  $E_{jc}$ , respectively, the increase of  $\tau_f$  can be calculated with sufficient accuracy by inserting the quantities derived above into the charge expressions given in section 3.5.1. However the associated storage times need to be calculated by including the derivatives of the respective widths, velocity etc.. The corresponding expressions can become quite complicated though and partially lead to discontinuities at  $I_{CK}$ . Hence, significant effort is necessary to bring the expressions into a form that is suitable for compact models. Beyond the critical current the charge expressions are no longer valid because the diffusion component in the electron current density  $J_{nx}$  is no longer negligible in the BC junction region.

### 3.5.3 High current densities

As discussed above the electric field at the BC junction decreases with current. Once the critical current is reached the electric field has become too small to support a drift current only, leading to an additional diffusion component. The required carrier density gradient in conjunction with the neutrality around the BC junction region causes an additional hole charge, which increases rapidly, i.e. over-proportional with current. Obviously, the electrical behavior of a transistor operated at high current densities is strongly determined by the effects occurring around the BC junction. This can be quite well observed in Fig. 2.12 where the collector and base storage time are the main contributors to the strong increase of the overall storage time. Therefore, these two storage times are discussed first before the other components are treated.

### 3.5.3.1 Collector region

Charge storage in the collector is caused by an elevated hole density within the injection region. The respective equations shall be derived below in order to obtain a feeling for their validity and the required assumptions. The derivation starts with that of the injection region, and is subdivided again into the case of low and of high voltages. Afterwards, the collector minority charge and associated storage time are derived.

#### A Low collector voltage

Figure 3.23 shows a sketch of the electric field distribution in the BC region that summarizes the results obtained from 1D device simulation for analytical calculations. The numbering of the curves corresponds to that in Fig. 3.4. Beyond the corresponding critical current  $I_{CKI}$  (curve 3) the injection zone starts to build (curve 4). In principle, the field within  $w_i$  is finite, but (much) smaller than in the remaining collector region, which is not conductivity modulated. The most simple approximation of the field distribution in this situation is shown as curve 4' (dash-dotted), in which the field within  $w_i$  is completely neglected and in the region  $w_{Ci} - w_i$  a simple *ohmic*-type voltage drop (i.e. constant field) is assumed. This assumption results from considering a lightly doped and fairly wide collector which is used for applications that need to sustain larger (breakdown) voltages such as in power amplifiers.

The voltage drop for curve 4' is simply given by

$$V_{ci} = I_T r_{Ci0} \frac{w_{Ci} - w_i}{w_{Ci}} \quad (3.260)$$

with  $r_{Ci0}$  from (3.243). Using  $I_{CK0}$  from (3.244), the current dependent injection width follows as

$$w_i = w_{Ci} \left( 1 - \frac{I_{CK0}}{I_T} \right), \quad (3.261)$$

which was first obtained in [10]. This equation is the result of a quite simplified distribution of the field in the collector and obviously valid only for fairly low voltages and quite high current densities. There have been various attempts in the past to derive an improved equation for the injection

width or a current dependent field description, but they turned out to be not feasible for compact modeling.

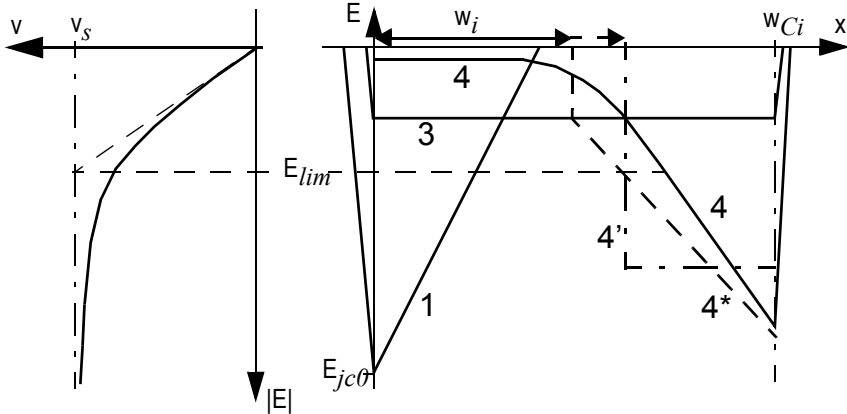


Fig. 3.23: Schematic electric field and carrier velocity at *low* voltages for  $N_{Ci}(x) = \text{const.}$

## B High collector voltage

Figure 3.24 shows a sketch of the electric field for important special cases, summarizing the results obtained earlier from 1D device simulation. The numbering of the curves corresponds to that in Fig. 3.2. For an analytical description of the injection zone, the distribution 5' is used (bold solid line), in which a slightly higher electric field ( $E(x) = -E_{lim}$ ) is assumed than in reality (bold dashed curve 5\*). The reason for this assumption is to ensure compatibility between the critical current formulation and the injection zone equation. Outside of the injection zone, the carriers possess saturation velocity, so that the slope of the electric field is given by the Poisson equation, and the field itself is given by  $(x' = x - x_{jc})$

$$E(x') = E(w_i) + \frac{qN_{Ci}}{\varepsilon} \left(1 - \frac{I_T}{I_{lim}}\right) x'. \quad (3.262)$$

Second integration equals the applied voltage

$$V_{ci} = E_{lim} w_{Ci} - \frac{qN_C}{2\varepsilon} \left(1 - \frac{I_T}{I_{lim}}\right) (w_{Ci} - w_i)^2 \quad (3.263)$$

where  $E(w_i) = -E_{lim}$  was inserted. Normalizing  $w_i$  to  $w_{Ci}$ , introducing  $V_{PT}$  and  $V_{lim}$ , and multiplying terms yields after division by  $V_{PT}$ :

$$\frac{V_{ci} - V_{lim}}{V_{PT}} + 1 - \frac{I_T}{I_{lim}} - 2\left(1 - \frac{I_T}{I_{lim}}\right) \frac{w_i}{w_{Ci}} + \left(1 - \frac{I_T}{I_{lim}}\right) \left(\frac{w_i}{w_{Ci}}\right)^2 = 0.$$

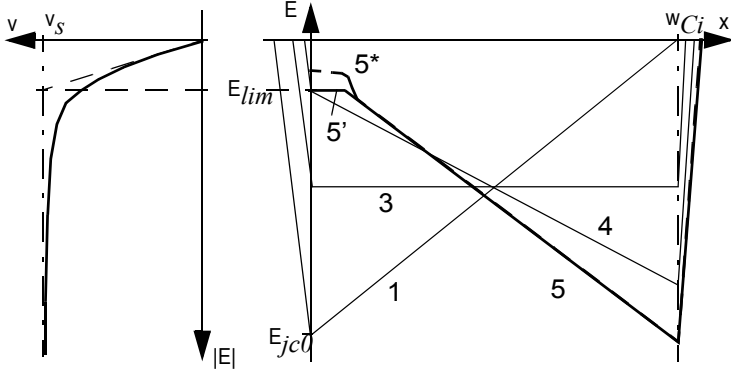


Fig. 3.24: Schematic electric field and carrier velocity at *high* voltages for  $N_{Ci}(x) = \text{const.}$

Replacing the first two terms by  $I_{CKh}/I_{lim}$  from (3.250), lumping the resulting first two terms together and dividing by  $(1 - I_T/I_{lim})$  leads to

$$\frac{I_{CKh} - I_T}{I_{lim} - I_T} - 2 \frac{w_i}{w_{Ci}} + \left(\frac{w_i}{w_{Ci}}\right)^2 = 0, \quad (3.264)$$

The solution of the quadratic equation then is

$$\frac{w_i}{w_{Ci}} = 1 - \sqrt{1 - \frac{I_{CKh} - I_T}{I_{lim} - I_T}} = 1 - \sqrt{\frac{I_{CKh}/I_{lim} - 1}{I_T/I_{lim} - 1}}. \quad (3.265)$$

Note that in [7]  $V_{lim} = 0$  is assumed in  $I_{CKh}$  from (3.250); i.e. the onset of high-current effects was defined there for  $E_{jc} = 0$ , which leads to a far too high value for the critical current.

Figure 3.25 shows as reference the results for the injection width from device simulation using the Regional Approach, which are compared to the derived analytical results. The low-voltage expression (3.261) gives reasonable results over the entire voltage range, while the high-voltage expression (3.265) can only be applied for  $I_T > I_{lim}$ , which is not fulfilled at lower voltages. The high voltage expression appears to match the shape of

the curve a little bit better for lower collector doping, but is worse for higher doping. Note that in (3.261)  $I_{CK0}$  was replaced by  $I_{CKI}$  with the same field dependent coefficient  $\alpha_m$  ( $= 1.11$ ) for the mobility as in DEVICE.  $I_{CKh}$  in (3.265) was calculated from (3.254), which already gives lower values around  $V_{lim}$  than [7]. From these results the simple expression (3.261) appears to be more suitable for compact modeling.

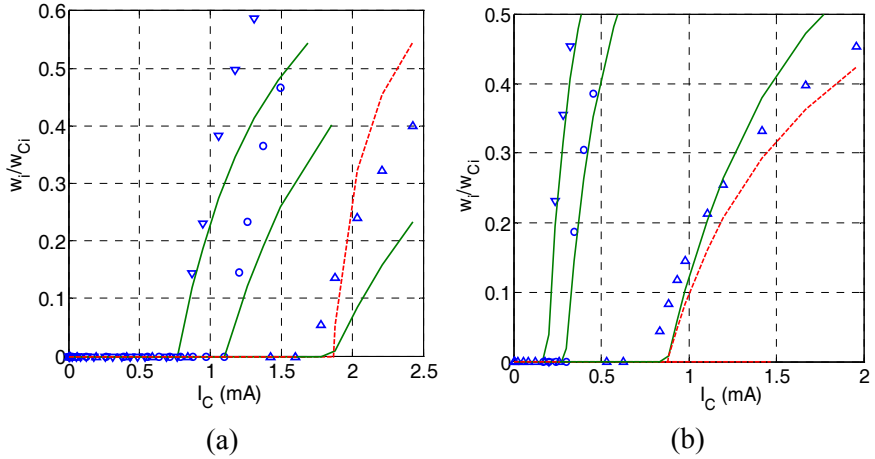


Fig. 3.25: Injection zone width (normalized to collector epi width) as function of collector current for the 1D BJT: Comparison of the analytical expressions (3.261) (solid line) and (3.265) (dashed line) with device simulation results obtained using REGAP (symbols). (a)  $N_{Ci} = 10^{17} \text{cm}^{-3}$  and  $V_{B'C'}/V = -2.3, 0, 0.4$ ; (b)  $N_{Ci} = 2 \cdot 10^{16} \text{cm}^{-3}$  and  $V_{B'C'}/V = -3.3, 0, 0.4$ .

### C Charge and storage time

The derivation below is based on Poisson's equation as well as both transport equations with the three unknowns  $p$ ,  $n$  and  $E_x$ . Under a minimum set of assumptions these equations are solved and the stored collector charge is then obtained as function of bias. The charge storage in the collector at high current densities was first analyzed in [45].

Since the gradient of the minority quasi-Fermi potential remains very small the net hole current density is still negligible,

$$J_{px} = -q\mu_p V_T \frac{dp}{dx} + q\mu_p p E_x \cong 0,$$

allowing to express the electric field as a function of carrier density:

$$E_x = \frac{V_T}{p} \frac{dp}{dx}. \quad (3.266)$$

Since the injection zone is electrically neutral,  $p = n - N_{Ci}$ . Hence, for the gradients of the carrier densities  $dn = dp$  holds. Inserting this and the field back into the *electron* transport equation gives

$$J_{nx} = 2q\mu_{nC0}V_T \left( 1 + \frac{N_{Ci}}{2p} \right) \frac{dp}{dx}. \quad (3.267)$$

Since  $J_{nx}$  is spatially constant, this equation can be integrated for  $x$  and  $p$ , yielding

$$x - x_{jc} = \frac{2q\mu_{nC0}V_T}{J_{nx}} \left[ p(x) - p(x_{jc}) + \frac{N_{Ci}}{2} \ln \left( \frac{p(x)}{p(x_{jc})} \right) \right], \quad (3.268)$$

which is an implicit relation for the spatial dependence of the hole density. The  $\ln()$  term in the brackets becomes significant only around  $x = w_i$  but is negligible over most of the injection zone. This is confirmed in Figs. 3.2 and 3.4, which reveal that the hole density decreases almost linearly from its value  $p(x_{jc})$  to a negligible value  $p(w_i)$ . Thus, once at sufficiently high injection  $p(x_{jc})/N_{Ci} \gg 1$  the corresponding term in (3.268) is negligible. Then, after replacing  $J_{nx}$  by  $-I_T/A_E$  one obtains the explicit relation [45]

$$p(x) = p(x_{jc}) - \frac{I_T}{2qA_E\mu_{nC0}V_T} (x - x_{jc}). \quad (3.269)$$

From this and noticing that  $p(w_i)$  is negligible, one obtains

$$p(x_{jc}) = \frac{I_T}{2qA_E\mu_{nC0}V_T} w_i. \quad (3.270)$$

as a function of transfer current and the spatial hole distribution

$$p(x) = \frac{I_T}{2qA_E\mu_{nC0}V_T} [w_i - (x - x_{jc})]. \quad (3.271)$$

In order to provide a feeling for the validity of the above assumptions, the bias dependence of  $p(x_{jc})$  is shown in Fig. 3.26a for different collector voltages. The analytical equation (3.270) is slightly shifted at high current densities, but otherwise captures the trend quite well. For comparison also



the electron density at the BC junction,  $n(x_{jc})$ , has been inserted to show how closely the two carrier densities are related, especially at high current densities. The linear increase of  $n(x_{jc})$  with current at lower injection results from the already existing drift current in the BC SCR.

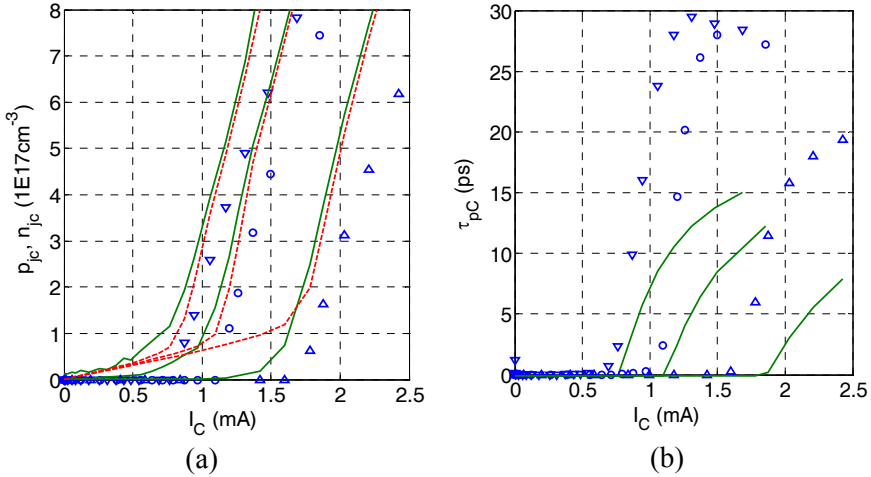


Fig. 3.26: (a) Carrier densities at the BC junction vs. collector current for the 1D BJT: comparison between hole density  $p(x_{jc})$  (solid lines), electron density  $n(x_{jc})$  (dashed lines) and  $p(x_{jc})$  from equation (3.270) (symbols).  $V_{B'C'}/V = 0.4, 0, -2.3$ . (b) Corresponding collector storage time vs. collector current: comparison between device simulation results (symbols) and analytical relation (3.273) (solid lines).

Integration over the injection region gives the stored excess collector hole charge

$$Q_{pC} = qA_E \int_{x_{jc}}^{x_{jc} + w_i} p dx = \frac{I_T}{4\mu_{nCi0}V_T} w_i^2. \quad (3.272)$$

The derivative yields the collector storage time

$$\tau_{pC} = \left. \frac{dQ_{pC}}{dI_T} \right|_{V_{CE}} = \frac{1}{4\mu_{nCi0}V_T} \left[ w_i^2 + 2I_T w_i \frac{dw_i}{dI_T} \right]. \quad (3.273)$$

In both equations an expression for the injection width as a function of bias has to be inserted which depends on the respective compact model. In order to assess the accuracy of the storage time expression, it is compared in

Fig. 3.26b to the results obtained from device simulation for different collector voltages. For the injection width the most simple expression (3.261) was employed (with  $I_{CKI}$  instead of  $I_{CKO}$ ) in order to be able to calculate the derivative. In contrast to older and power transistors (with lower doping concentration) the collector storage time increases much more rapidly at the onset of high-current effects and can be roughly described by the analytical equation using technology and physics-based parameters only. Note that also the saturation time constant

$$\tau_{pCs} = \frac{w_{Ci}^2}{4\mu_{nCi0}V_T}, \quad (3.274)$$

which is reached once the injection zone hits the buried layer, appears smaller than in device simulation. These deviations are partially caused by the impact of high-doping effects (i.e. bandgap-narrowing) in the base, which can form energy barriers at the BC junction. Also note that the high-voltage solution (3.265) only works at the highest voltage in Fig. 3.26b but is still less accurate.

The formation of a neutral zone  $w_i$  in the collector is often called *base widening* [46] in the classical literature, a designation that was coined first in [10]. Unfortunately, this designation generates an incorrect perception of the actually happening mechanisms and, as a result, has caused a variety of physically incorrect charge storage models. The most common mistake is to simply replace  $w_B$  with  $(w_{Bm} + w_i)$  in the base transit time (e.g. [47, 48, 17] and SGPM extensions in various circuit simulators). The reasons why the base widening approach leads to an incorrect model are:

- The model equation results from the definition  $\tau = Q/I$  rather than from  $\tau = dQ/dI$  as required from small-signal measurements.
- The assumed quadratic dependence of  $\tau_f$  with  $w_B$  is valid only for the *base* region and a *diffusion* transistor.
- The material properties (e.g. mobility, bandgap) in the collector region are different from those in the base region but employed in the incorrect definition.

### 3.5.3.2 Neutral base region

At the beginning of the high-current region, i.e.  $I_T = I_{CK}$ , the collector-sided boundary of the neutral base ( $x_c$ ) reaches the BC junction, but the minority carrier density in the base region is still small compared to the doping density. As long as this is the case the charge is still given by (3.238), and the same thoughts as for medium current densities still hold. Therefore,  $\tau_{Bfd}$  and  $\tau_{Bfv}$  are given by (3.241) and (3.242), where the field derivative terms now start to dominate the storage times.

Once the Webster-effect [15] sets in at very high current densities, the doping related portion of the built-in field that acts upon the minority carriers decreases to its bandgap-narrowing related value  $E_{bg}$  as was discussed during the derivation of the transfer current. In the limit of very high injection, this leads to a lower drift factor  $\zeta_h \approx |E_{bg}|/(V_T/w_{Bm})$  instead of  $\zeta$  and a higher average mobility  $\bar{\mu}_{nBh}$ . The spatial dependence of the electron density is now given by (3.73), which is repeated here with  $w_B = w_{Bm}$ :

$$n(x) = \frac{I_T w_{Bm}}{2qA_E \bar{\mu}_{nBh} V_T \zeta_h} \left[ 1 - \exp\left(-\zeta_h \left(1 - \frac{x-x_c}{w_{Bm}}\right)\right) \right] + \frac{I_T w_{Ci}}{2qA_E \bar{\mu}_{nCi0} V_T} \exp\left(-\zeta_h \left(1 - \frac{x-x_c}{w_{Bm}}\right)\right).$$

Also,  $n_c = n(x_{jc}) \approx p(x_{jc})$  was inserted from (3.270) with  $w_{Ci}$  replacing  $w_i$ , since for most transistor designs the injection zone reaches the epi width at current densities lower than those for base conductivity modulation. Note, that the boundary condition  $n_c = I_T/(qA_E v_{sn})$  is no longer valid since the current density now contains also a large diffusion component. Integration of  $n(x)$  yields the neutral base charge

$$Q_{nB} = \left[ \frac{w_{Bm}^2}{2\bar{\mu}_{nBh} V_T F_{\zeta_h}} + \frac{w_{Ci} w_{Bm}}{2\bar{\mu}_{nCi0} V_T G_{\zeta_h}} \right] I_T \quad (3.275)$$

with the limiting values for the factors

$$F_{\zeta_h} = \frac{\zeta_h^2 f_{\zeta_h}}{(\zeta_h - 1)f_{\zeta_h} + 1} \quad \text{and} \quad G_{\zeta_h} = \frac{\zeta_h f_{\zeta_h}}{f_{\zeta_h} - 1}, \quad (3.276)$$

which do not depend on bias anymore. Notice, that at (very) high-injection the base charge again depends only linearly on current, as is observed in 1D device simulation.

Finally, the derivative of the base charge gives the saturation values for the drift-diffusion storage time,

$$\tau_{Bfds} = \frac{w_{Bm}^2}{2\bar{\mu}_{nBh}V_T F_{\zeta h}}, \quad (3.277)$$

and for the carrier jam storage time,

$$\tau_{Bfvs} = \frac{w_{Ci}w_{Bm}}{2\mu_{nCi0}V_T G_{\zeta h}}. \quad (3.278)$$

Comparison to the low-current expressions reveals that the collector velocity  $v_{cn}$  has now been replaced by  $\mu_{nCi0}/(2V_T/w_{Ci})$  in which the electric field  $2V_T/w_{Ci}$  is consistent with the result for the field obtained from the derivative of the quasi-Fermi potential (as the driving force) towards infinite current (cf. sec. 4.4.2). For negligible bandgap-variation  $F_{\zeta} = 2$ , resulting in the classical expression for the high-injection base transit time,

$$\tau_{Bfds} = \frac{w_{Bm}^2}{4\bar{\mu}_{nBh}V_T}. \quad (3.279)$$

In the transition region between medium and high injection, equation (3.238) can be employed as first order approximation, but now with a bias dependence included in the mobility as well as the factors  $F_{\zeta}$  and  $G_{\zeta}$  from (3.172) and (3.175), respectively. Note, that towards very high injection, the derivatives w.r.t.  $E_{jc}$  in (3.241) and (3.242) drop back to zero. This can lead to a “spike” in the (1D) current dependence of the storage time as often observed in device simulation. Usually, however, the current density at which such a spike occurs is far too high for circuit applications.

Classical transistor theory assumes  $\zeta_h = 0$  and separation into a charge component for “forward” mode of operation,  $Q_{nBf}$  and one for inverse (or reverse) mode of operation,  $Q_{nBr}$ . This is equivalent to a superposition of two linearly independent carrier density distributions controlled by the respective voltage across the junctions. However, this is not a valid concept for practical transistor structures since the electron density in the collector region is modulated by the BE voltage.

### 3.5.3.3 BC space-charge region

Once  $I_T$  has reached  $I_{CK}$  and an injection region forms, the BC SCR has disappeared at  $x_{jc}$ . However, as Figs. 3.2 and 3.4 show now a current induced SCR starts to form in the collector region "behind" the injection zone (i.e. in  $x = [w_i, w_{Ci}]$ ). For very low BC voltages first still an ohmic region remains but since the region  $(w_{Ci} - w_i)$  shrinks with increasing current density, eventually the field and carrier density difference  $(n - N_{Ci})$  become large enough again to form a SCR. The corresponding storage time shall be derived next. Again, as for low current densities, small-signal variables are used (indicated by  $\delta$ ). Also, the bias dependent charge stored within the current induced SCR is now given by

$$\delta Q_{BC} = A_E \int_{w_i}^{w_{Ci}} \delta \rho(x) dx. \quad (3.280)$$

Integrating Poisson's equation once from  $x = w_i$  to an arbitrary location  $x$  within the SCR to the end of the collector gives for the electric field variation

$$\delta E(x) - \delta E(w_i) = \frac{1}{\epsilon} \int_{w_i}^x \delta \rho(\xi) d\xi \quad (3.281)$$

with an arbitrary spatial dependence of the small-signal space charge  $\delta \rho$  and assuming a spatially independent permittivity. At high current densities the electric field at  $w_i$  is negligible. Integrating again from  $w_i$  to the end of the collector and assuming a negligible extension of the SCR into the buried-layer then gives the voltage across the SCR

$$\delta V_{BC} = - \int_{w_i}^{w_{Ci}} \delta E(x) dx. \quad (3.282)$$

Inserting  $\delta E(x)$  from (3.281) leads to

$$\epsilon \delta V_{BC} = - \int_{w_i}^{w_{Ci}} \left( \int_x^{w_{Ci}} \delta \rho(\xi) d\xi \right) dx. \quad (3.283)$$

which has the same form as (3.190) except for the integration limits. Following similar partial integration steps as in section 3.5.1.2 as well as replacing  $w_{BC}$  by  $w_{Ci}$  and  $x = 0$  by  $x = w_i$  yields to one expression that is similar to (3.193) and another one that reads

$$\left( x \int_x^{w_{Ci}} \delta \rho(\xi) d\xi \right) \Big|_{w_i}^{w_{Ci}} = -w_i \int_{w_i}^{w_{Ci}} \delta \rho(\xi) d\xi = -w_i \frac{\delta Q_{BC}}{A_E}, \quad (3.284)$$

where the charge from (3.280) has been used for the last term. Inserting the expressions found above with into (3.282) gives after some rearrangement

$$\varepsilon \delta V_{B'C} = w_i \frac{\delta Q_{BC}}{A_E} + \left( w_{BC} \frac{\delta Q_{BC}}{A_E} - w_{BC} \int_{w_i}^{w_{Ci}} \left( 1 - \frac{x}{w_{BC}} \right) \delta \rho(x) dx \right).$$

With  $(w_i + w_{BC}) = w_{Ci}$  the BC charge follows as

$$\delta Q_{BC} = \frac{\varepsilon \delta V_{B'C}}{w_{Ci}} A_E + \frac{w_{BC}}{w_{Ci}} A_E \int_{w_i}^{w_{Ci}} \left( 1 - \frac{x}{w_{BC}} \right) \delta \rho(x) dx. \quad (3.285)$$

The first term corresponds to the punch-through capacitance

$$C_{jCPT} = A_E (\varepsilon / w_{Ci}), \quad (3.286)$$

while the second term is current dependent.

The equivalent of the above electron charge variation  $dQ_{BC}$  caused by a change in  $V_{B'C}$  or  $I_T$  (within  $\rho$ ) must be compensated by a hole charge variation,  $dQ_{p,BC}$ , on the base side or within the injection zone. As a consequence, the contribution of the SCR related time constant to the transit frequency (defined in common-emitter configuration) reads

$$\left. \frac{dQ_{p,BC}}{dI_T} \right|_{V_{CE}} = -C_{jCPT} g_{mi} + \tau_{BCvh} \quad (3.287)$$

with the time constant

$$\tau_{BCvh} = -\frac{w_{BC}}{w_{Ci}} A_E \int_{w_i}^{w_{Ci}} \left( 1 - \frac{x}{w_{BC}} \right) \frac{d\rho(x)}{dI_T} \Big|_{V_{CE}} dx. \quad (3.288)$$

In order to evaluate the integral term a specific expression for the space-charge has to be inserted. Assuming the most simple case of a constant doping profile and average carrier velocity  $v_c$  in the SCR gives with  $I_T = q A_E n v_c$  the bias dependent expression

$$\rho(x) = q(N_{Ci} - n) = qN_{Ci} - \frac{I_T}{A_E v_c} < 0. \quad (3.289)$$

This results in the high-current region time constant

$$\tau_{BCvh} = \frac{w_{BC} w_{BC}}{w_{Ci} 2v_c}. \quad (3.290)$$

Compared to (3.201), the weight factor  $w_{BC}/w_{Ci}$  leads to a smaller than expected value that also decreases faster with increasing current. It should be noted though that the current densities considered in above derivation are usually far too high for practical circuit applications. Also, other time constants dominate.

### 3.5.3.4 Neutral emitter region

For conventional transistor structures (high emitter doping) the low-injection case applies under all relevant bias conditions. As a consequence, the bias dependence of the emitter storage time is determined only by the small-signal current gain. Therefore, the same considerations are valid that led to the description (3.257) to (3.259) for medium current densities. In fact the assumptions made in that derivation are now even more valid.

### 3.5.3.5 BE space-charge region

The formation of an injection zone and the associated strong increase in the hole charge leads to a significant reduction in the *increase* of both  $I_T$  and transconductance  $g_{mi}$ . If the proportionality of  $I_T$  with  $V_{B'E'}$  is represented by a (bias dependent) non-ideality coefficient  $m_C$ , then the value of the latter starts from close to 1 at low current densities, then approaches and passes 2 at high current densities beyond  $I_{CK}$ , and eventually approaches again 2 at *very* high current densities when base conductivity modulation has set in. A value of 2 eliminates the current dependence of the term  $g_{mi}^{-1} \exp(v_{B'E'}/2V_T)$  in (3.234), and a value larger than 2 leads even to an increase of  $\tau_{BE}$  with current which can be observed in Fig. 3.20.

However, since at *very* high current densities the BE SCR eventually disappears  $Q_{BE}$  will also disappear, and the charge contribution close to the junction will ultimately be absorbed in the neutral base and emitter

storage charge components. The associated current densities though are way beyond any practical usefulness.

### 3.5.3.6 Summary

The mechanisms described in this and the previous section are often called *high-current effects* and include the so-called Kirk and *quasi-saturation effect* (forward biased internal BC junction for reverse bias at external BC terminals). Some of the derived expressions for the charge and storage time are not compatible with those for low or medium current densities. It is the task of compact modeling to make the appropriate compromises and introduce adequate smoothing functions in order to obtain expressions that are suitable for compact models.

## 3.6 References

- [1] J. Lindmayer and C.Y. Wrigley, "Fundamentals of Semiconductor Devices", D. Van Nostrand Comp. Inc., Toronto, 1965.
- [2] R.L. Pritchard, "Electrical Characteristics of Transistors", McGraw-Hill, New York, 1967.
- [3] A.B. Phillips, "Transistor Engineering", McGraw Hill, New York, 1962.
- [4] C.-T. Sah, "Fundamentals of Solid-State Electronics", World Scientific, Singapore, 1991.
- [5] Y. Taur and T.H. Ning, "Fundamentals of modern VLSI devices", Cambridge Univ. Press, 1998.
- [6] M. Schröter and H.-M. Rein "Transit Time of High-Speed Bipolar Transistors in Dependence on Operating Point, Technological Parameters, and Temperature", Proc. IEEE Bipolar Circuits and Technology Meeting (BCTM), Minneapolis, pp. 250-253, 1989.
- [7] R.J. Whittier and D.A. Tremere, "Current Gain and Cutoff Frequency Falloff at High Current Densities", IEEE Trans. Electron Dev., Vol. ED-16, pp. 39-57, 1969.
- [8] D.L. Bowler and F.A. Lindholm, "High Current Regimes in Transistor Collector Regions", IEEE Trans. Electron Dev., Vol. ED-20, pp. 257-263, 1973.



- [9] K. Lee, A.-H. Cho, K.-W. Park, B. Kim, "Improved VBIC model for SiGe HBTs with an unified model of heterojunction barrier effects", *IEEE Trans. Electron Dev.*, Vol. 53, No. 4, pp. 743-752, 2006.
- [10] C.T. Kirk, "A Theory of Transistor Cutoff Frequency (f<sub>T</sub>) Falloff at High Current Densities", *IEEE Trans. Electron Dev.*, Vol. ED-9, pp. 164-174, 1962.
- [11] H. Rüchardt, "Effects in mesa transistors at high current densities" (in German), *NTZ*, Vol. 15, pp. 333-340, 1962.
- [12] M. Schröter, H. Tran, "Modeling of charge and collector field in Si-based bipolar transistors", *Proc. WCM, International NanoTech Meeting*, Boston (MA), pp. 102-107, 2004.
- [13] L. Turgeon and J. Matthews, "A bipolar transistor model of quasi-saturation for use in computer aided design", *IEDM Tech. Dig.*, pp. 394-397, 1980.
- [14] G.M. Kull et al., "A Unified Circuit Model for Bipolar Transistors Including Quasi-Saturation Effects", *IEEE Trans. Electron Dev.*, Vol. ED-32, pp. 1103-1113, 1985.
- [15] W.M. Webster, "On the Variation of Junction-Transistor Current-Amplification Factor with Emitter Current", *Proc. IRE*, Vol. 42, pp. 914-920, 1954.
- [16] C. McAndrew, J.A. Seitchik, D.F. Bowers, M. Dunn, M. Foisy, I. Getreu, M. McSwain, S. Moinian, J. Parker, D.J. Roulston, M. Schröter, P. v. Wijnen and L.F. Wagner, "VBIC95, the vertical bipolar intercompany model", *IEEE J. Solid-State Circuits*, Vol. 31, pp. 1476-1483, 1996.
- [17] P. Antognetti and G. Massobrio, "Semiconductor Device Modeling with SPICE", McGraw-Hill, 1988.
- [18] H.C. de Graaff and W.J. Kloosterman, "New Formulation of the Current and Charge Relations in Bipolar Transistor Modeling for CACD Purposes", *IEEE Trans. Electron. Dev.*, Vol ED-32, pp. 2415-2419, 1985.
- [19] J.W. Slotboom and H.C. De Graaff, "Bandgap Narrowing in Silicon Bipolar Transistors", *IEEE Trans. Electron Devices*, vol. ED-24, pp. 1123-1125, 1977.
- [20] W. Shockley and W. T. Read, "Statistics of recombination of holes and electrons", *Phys. Rev.* Vol. 87, pp. 835-842, 1952.
- [21] R.N. Hall, "Electron-hole recombination in Germanium", *Phys. Rev.*, Vol. 87, p.387, 1952.

- [22] C. T.-Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in p-n junction and p-n junction characteristics", *Proc. IRE*, Vol. 45, pp. 1228-1248, 1957.
- [23] V. Kunz, C. deGroot, S. Hall, and P. Ashburn., "Polycrystalline Silicon-Germanium emitters for gain control, with application to SiGe HBTs", *IEEE Trans. Electron Dev.*, Vol. 50, pp. 1480-1486, 2003.
- [24] K. Seeger, "Semiconductor physics - An introduction", Springer Series in Solid-State Sciences, Springer Verlag Heidelberg, pp. 319, 1997.
- [25] S.M. Sze, "Physics of Semiconductor Devices", Wiley & Sons, New York, 1981.
- [26] K. Thornber, "Applications of scaling to problems in high-field electronic transport", *J. Appl. Phys.*, Vol 52, pp. 279-290, 1981.
- [27] W. Shockley, "Problems related to p-n junctions in silicon", *Czech J. Physics*, Vol. B11, pp. 81-122, 1961.
- [28] W.J. Kloosterman and H.C. de Graaff, "Avalanche Multiplication in a Compact Bipolar Transistor Model for Circuit Simulation", *IEEE Trans. Electron Dev.*, Vol. ED-36, pp. 1376-1380, 1989.
- [29] M. Schröter, Z. Yan, T.-Y Lee and W. Shi, "A compact tunneling current and collector breakdown model", *Proc. IEEE Bipolar Circuits and Technology Meeting*, Minneapolis, pp. 203-206, 1998.
- [30] S. Miller, "Avalanche breakdown in Germanium", *Phys. Rev.*, Vol. 99, pp. 1234-1241, 1955.
- [31] H.K. Gummel and D.L. Scharfetter, "Depletion layer capacitance of p+n step junctions", *J. Appl. Phys.*, Vol. 38, No. 5, pp. 2148-2153, 1967.
- [32] B.R. Chawla and H.K. Gummel, "Transition Region Capacitance of Diffused p-n Junctions", *IEEE Trans. Electron Dev.*, Vol. ED-18, pp. 178-187, 1971.
- [33] C.Y. Kwok, "Space charge and injection capacitances of pn junctions from small signal numerical solutions", *J. Appl. Phys.*, Vol. 16, pp. 263-267, 1983.
- [34] H.F. Chau et al., "Breakdown-speed considerations in AlGaAs/GaAs heterojunction bipolar transistors with special collector design", *IEEE. Trans. Eletron Dev.*, Vol. ED-39, pp. 2711-2719, 1992.
- [35] H. Krömer, "Two-Integral equations pertaining to the electron transport through a bipolar transistor with nonuniform energy gap in

- the base region", *Solid-State Electron.*, Vol. 28, pp. 1101-1103, 1985.
- [36] K. Suzuki, "Emitter and base transit time of polycrystalline silicon emitter contact bipolar transistors", *IEEE Trans. Electron Dev.*, Vol. ED-38, pp. 2512-2518, 1991.
- [37] H. Krömer, "Zur Theorie des Diffusions- and des Drifttransistors", *Archiv d. elektr. Übertragung (AEÜ)*, Vol. 8, pp. 223-228, pp. 363-369, pp. 499-504, 1954.
- [38] M. Schröter and T.-Y. Lee, "A physics-based minority charge and transit time model for bipolar transistors", *IEEE Trans. Electron Dev.*, vol. 46, pp. 288-300, 1999.
- [39] M. Rohner, "Physical limitations of InP/InGaAs heterojunction bipolar transistors", *Diss. ETH Zurich*, Vol. 125, Hartung Gorre Konstanz.
- [40] R.G. Meyer and R.S. Muller, "Charge-Control Analysis of the Collector-Base Space-Charge-Region Contribution to Bipolar-Transistor Time Constant", *IEEE Trans. Electron Dev.*, Vol. ED-34, pp. 450-452, 1987.
- [41] M. Schröter, "Physical models for high-speed silicon bipolar transistors - a comparing overview", *Habilitation thesis (in German)*, Ruhr-University Bochum, 1994.
- [42] S.P. Morgan und F.M. Smits, "Potential Distribution and Capacitance of a Graded pn-Junction", *BSTJ*, Vol. 39, S. 1573-1602, 1960.
- [43] M. Malorny, M. Schröter, D. Celi, D. Berger, "An improved method for determining the transit time of Si/SiGe bipolar transistors", *Proc. BCTM*, pp. 229-232, 2003.
- [44] M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors with special regard to high current densities and two-dimensional effects", *Ph. D. thesis (in German)*, Ruhr-University Bochum, Bochum, Germany, 1988.
- [45] J.R. Beale and J.A. Slatter, "The Equivalent Circuit of a Transistor with a Lightly Doped Collector Operating in Saturation", *Solid-State Electronics*, Vol. 11, pp. 241-252, 1968.
- [46] F.M. Klaassen and H.C. de Graaff, "Compact Transistor Modelling for Circuit Design", *Springer-Verlag Wien New York*, 1990.
- [47] H.K. Gummel, "A Charge-Control Relation for Bipolar Transistors", *BSTJ* Vol. 49, pp. 115-120, 1970.

- [48] I.E. Getreu, “Modeling the Bipolar Transistor”, Amsterdam: Elsevier Scient. Publ. Comp., 1978.

## **Chapter 4**

### **Advanced Theory**

This chapter presents the qualitative and theoretical foundation for modeling SiGe HBTs. Many of the discussions go beyond the conventional textbook theory for bipolar transistors. Section 4.1 provides a basic understanding of the impact of the heterojunctions on current and charge in a transistor and its various regions. From this knowledge extensions of classical (BJT) theory can be derived. This starts in section 4.3 with the DC case and continues in section 4.4 with the charge storage for the dynamic case. Section 4.5 explains non-quasi-static (NQS) effects, which are present in both BJTs and HBTs, and provides an overview on possible options for the theoretical description along with the consequences for building compact model equivalent circuits. All considerations are still based on 1D structures.

## 4.1 HBT operation principle

The combination of semiconductors with different material properties within a device region provides additional degrees of freedom for device design to achieve better performance. However, although such compound devices benefit from the material composition, the latter often also introduces physical effects, which may be undesired and may impact device characteristics in certain bias or temperature regions. In this chapter the benefits and consequences of incorporating Ge into Si for fabricating SiGe HBTs are discussed from an electrical point of view. For fabrication related issues the reader is referred to [1, ch. 3].

The different arrangements of a heterojunction in a semiconductor region will be discussed step by step, starting in section 4.1.1 with a simple homogeneously doped semiconductor. Next, in section 4.1.2, a pn junction that resembles the BC junction of an npn HBT will be considered. In all cases, the formation of the respective (energy) band diagram is discussed in detail. Finally, in section 4.1.3 important internal variables of the HBT are presented that are relevant for the subsequent sections on theory development.

### 4.1.1 The SiGe heterojunction in a neutral region

As an illustrative example a simple 1D structure with uniform doping (i.e. spatially constant dopant impurity concentration) is considered. However, the structure is subdivided into a SiGe material portion with, e.g., 30% Ge mole fraction and a Si material region. The schematic (energy) band diagram of the two regions not in electrical contact is shown in Fig. 4.1a. The electron affinity in Si is slightly larger than in SiGe, but the energy gap in Si may be significantly larger than in SiGe. The Fermi level is shown simultaneously for a p- or n-doped sample. Its relative position to the valence and conduction band edge, respectively, is the same in both material regions since it is assumed that there is no difference in effective mass.

Figure 4.1b shows the energy diagram *immediately* after bringing the two regions together. *Immediately* means that restoring the charge balance has *not* yet taken place. Without applying any voltage at the contacts there

is no current flowing, and the Fermi level must remain flat throughout the structure. As a consequence, unless charge rebalancing occurs the distance between the Fermi level and the majority band edges differs. However, since the doping density remains the same, the band edges have to move back to their original distance to the Fermi level. This can only happen if carriers start moving to the respective lower energy levels, causing a space charge region and the electrostatic potential to change.

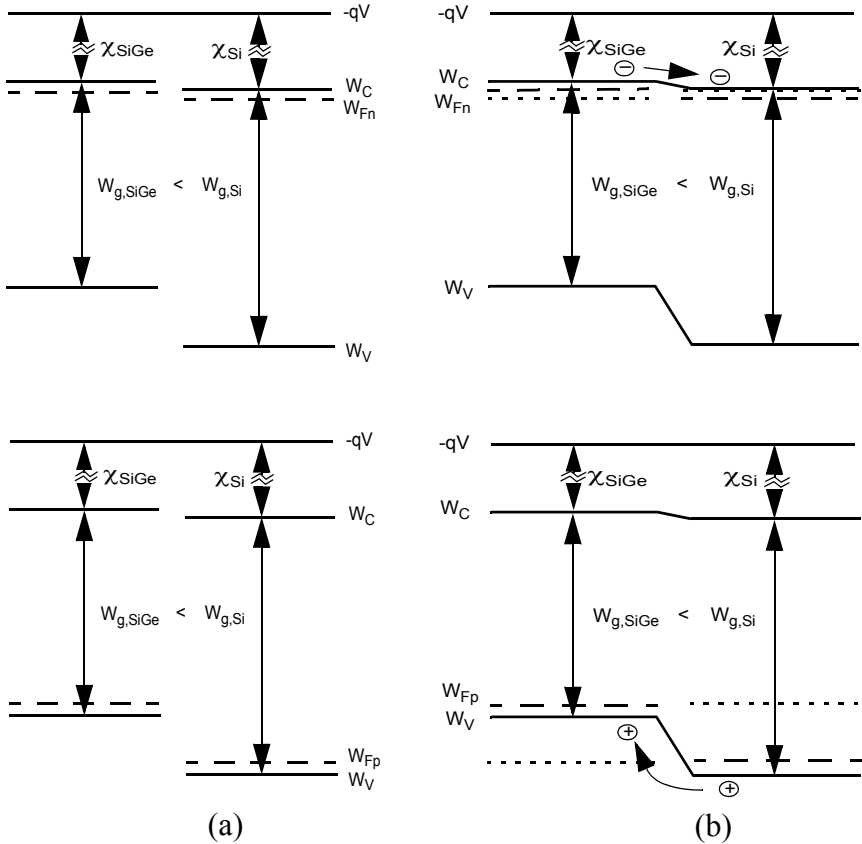


Fig. 4.1: Band diagram with (macro-)potential  $V$  for a uniformly n-doped structure (upper figure) and p-doped structure (lower figure) consisting of a SiGe portion on the r.h.s and a Si portion on l.h.s.: (a) Separate band diagrams for each material and (b) band diagram immediately after contacting without charge rebalancing.

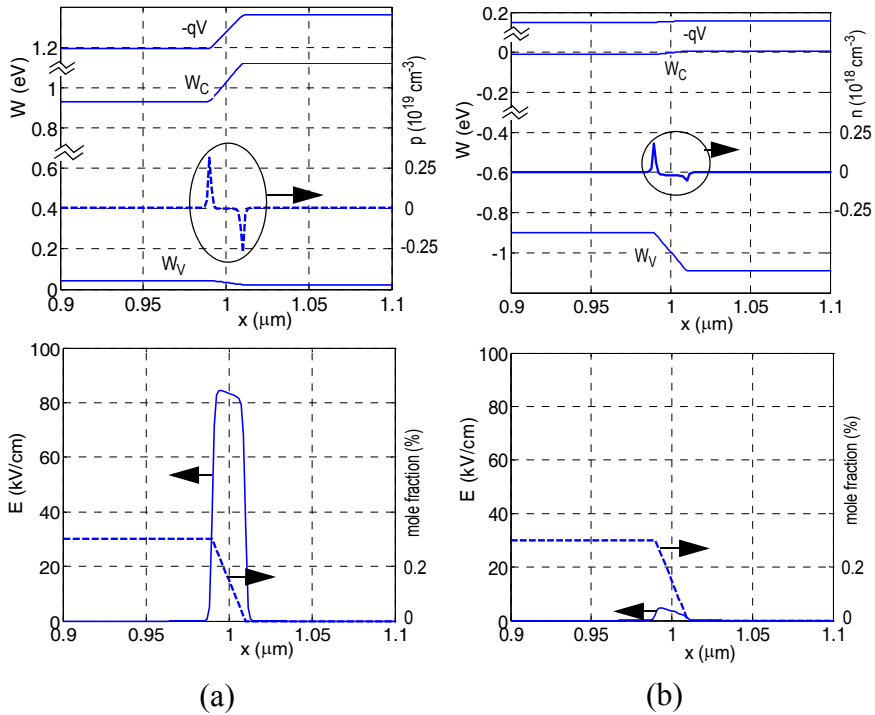


Fig. 4.2: Equilibrium band diagram with (macro-)potential and carrier density (upper figures) and electric field with mole fraction (lower figures) for a simple resistance structure with (a)  $p$  doping and (b)  $n$  doping of  $4 \cdot 10^{19} \text{ cm}^{-3}$ . The macro potential distance to the conduction band is not drawn to scale and was obtained from the simulated electrostatic potential by arbitrarily shifting the latter in order to move the curves graphically above the conduction band in the band diagram. Note the factor 10 difference in the hole density scale between (a) and (b).

The band diagram in equilibrium *after restoring the balance* depends on the type of doping. The following discussion is based on a quantitative example of a simple resistor with a length of  $2 \mu\text{m}$ , uniform doping of  $4 \cdot 10^{19} \text{ cm}^{-3}$ , and a steep linear Ge mole fraction drop in the middle between  $0.99$  and  $1.01 \mu\text{m}$  as shown in Fig. 4.2 (lower portion). Device simulation provides an exact solution of Poisson's equation after the balance has been restored. The corresponding band edges, macro-/electrostatic potential, majority carrier density, and electric field as a function of position are also displayed in Fig. 4.2. The two cases of positive and negative dop-



ing represent the neutral region of the base and collector or emitter, respectively, in an HBT and, hence, are important for understanding the electrical behavior of these regions under certain bias conditions.

First, the p-doped example is discussed. In this case, as indicated in Fig. 4.1b, holes will move from the Si region to the lower energy levels in the SiGe region leaving behind negatively charged acceptor atoms in Si. The resulting uncompensated space charge on either side leads to a positive electric field, that starts to pull back holes into the Si region. Once the drift current equals the diffusion current the balance between both sides has been restored. The field causes an electrostatic potential drop across this dipole layer which will completely compensate the previous step in the valence band in order to ensure charge neutrality (i.e.  $p = N_A$ ) in the remaining ohmic regions. Since the electrostatic potential is proportional to the macro-potential and the band edges they all align in parallel as shown in Fig. 4.2a. The slight remaining difference in  $W_V$  is caused by a difference in effective mass due to its mole fraction dependence that is taken into account in device simulation. The important point though is that the original valence band step, which occurred right after contacting in Fig. 4.1b, now appears in the conduction band impeding a possible flow of electrons (i.e. minority carriers) from the SiGe to the Si region.

The result for the n-doped sample shown in Fig. 4.2b can be explained similarly. In this case, the lower electron affinity in SiGe causes electrons to move to the lower energy levels in Si (cf. Fig. 4.1b). Again, this creates a dipole layer but with much smaller charge density as for the p-doped sample due to the much smaller energy difference in  $W_C$ . Nevertheless, a small positive electric field as well as the corresponding voltage drop are visible (cf. Fig. 4.2b). The associated difference in electrostatic potential equals the electron affinity difference and leads to the observed change in band edges. Again, the slight remaining difference is caused by the mole fraction dependent effective mass. It is important to note, that now the large energy step remains in the valence band, impeding a possible hole flow from the SiGe to the Si region.

#### 4.1.2 Heterojunction diode

The second example is a simple 1D pn junction that is uniformly doped on each side. Like in a transistor BC junction, the  $p$  side on the left in Fig. 4.3a is much higher doped than the  $n$  side on the right. Again, the structure is subdivided into a SiGe material portion with 30% Ge mole fraction and a Si portion as shown in Fig. 4.3b,c, with an abrupt transition between the two material regions. There are now three options, that need to be considered for the Ge drop location  $x_h$ , namely

- in the  $p$  region ( $x_h < x_j$ ),
- in the  $n$  region ( $x_h > x_j$ ), or
- at the junction ( $x_h = x_j$ ).

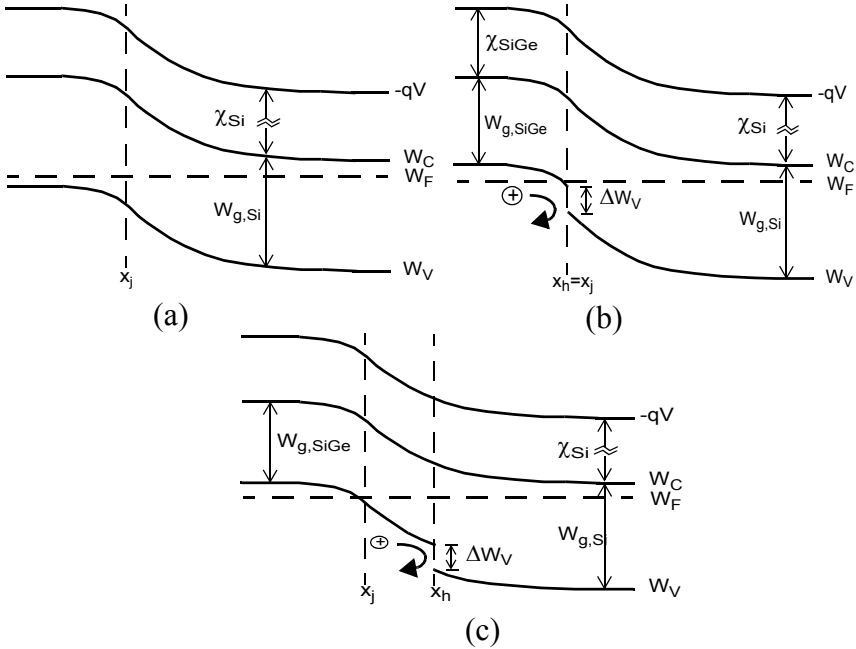


Fig. 4.3: Equilibrium band diagram with (macro-)potential for a 1D diode structure with high  $p$  doping on the left and low  $n$  doping on the right. (a) Homojunction. (b) Heterojunction with Ge drop (at  $x_h$ ) coinciding with the junction and (c) heterojunction with Ge drop moved into the SCR of the  $n$  region, each *immediately* after contacting the different regions.

As long as the heterojunction is located in the neutral regions, the results of the previous section apply. Once the heterojunction moves into the SCR, the effects from the material composition change are superimposed with those related to the space-charge occurring in a pn homojunction. Below, only the cases  $x_h \geq x_j$  will be discussed, since a Ge drop in the base region imposes a parasitic barrier that makes the transistor more or less useless [2,3].

The first step toward constructing the energy diagram is again to draw the band edges *immediately* after “turning on” the Ge composition, i.e. without taking into account charge movements. This is shown in Fig. 4.3b and c, which are the basis for the subsequent qualitative discussion. For the sake of simplicity it is assumed that electron affinity, effective mass, and permittivity do not change with composition and high doping effects are negligible.

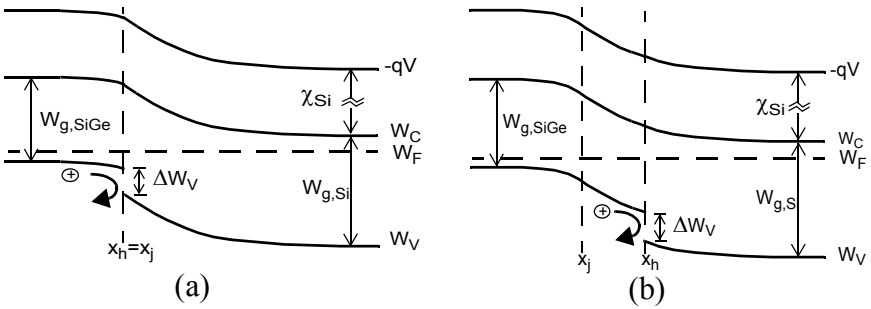


Fig. 4.4: Equilibrium band diagram with (macro-)potential for the 1D pn diode structure from Fig. 4.3 after restoring the balance through mobile carriers: heterojunction with (a)  $x_h = x_j$  and (b)  $x_h > x_j$ .

In Fig. 4.3b, the Ge drop at the junction leads to a discontinuity in the valence band at  $x_j$  (due to the band gap reduction in the SiGe p region), which acts like a barrier for holes that move to the right. Since in the snapshot shown the potential bending of the homojunction still exists, the valence band on the SiGe side has moved above the Fermi level. This is of course inconsistent with the equilibrium condition outside of the SCR where the distance between Fermi level and valence band edge is given by the doping. This condition can only be achieved by a reduction in the band bending within the SCR on both sides. As a consequence of a smaller band

bending, there must be also a smaller charge in the SCR. Since the charge in the SCR is given by the ionized impurity atoms, a smaller number of the latter must exist on both sides. In other words, compared to the homojunction case, a smaller number of mobile carriers crosses the heterojunction to establish equilibrium. Since the energy band bending must be the same at all locations, where bandgap, electron affinity, and effective mass do not change, the (original) barrier in the valence band now also influences the conduction band and macro-potential. The final result of this adjustment is shown in Fig. 4.4a.

The band edge adjustment for the case  $x_h > x_j$  occurs in a similar way. The barrier in the valence band prevents holes from diffusing into the n region, thus lowering the bending of the valence band. As a result, also the conduction band edge as well as the macro-potential show a smaller bending (cf. Fig. 4.4b).

It is interesting to note that the smaller conduction band bending reduces the diffusion barrier for electrons as compared to the homojunction. Therefore, more electrons will move into the p region and result in an increased electron density there. This is consistent with theory in which a smaller bandgap leads to a larger minority carrier density.

To aid the understanding of heterojunctions and provide a link to compact modeling, the (vertical) electric field and the carrier densities obtained from device simulation along with the doping concentration are shown in Fig. 4.5a, b. The homojunction results inserted for comparison. In all cases, the expected triangular shape for the field can be observed. In the SCR of the heterojunction the field peak is reduced due to the smaller ionized charge. Also, in case of  $x_h > x_j$ , the discontinuity in the permittivity, which is a function of material composition, causes the field to jump at the heterojunction. The carrier densities show the expected discontinuity at the heterojunction due to the bandgap change. Also, the electron density is larger (as minority carrier density) in the SiGe p-region than in the Si p-region (homojunction case) due to the bandgap reduction caused by the Ge.

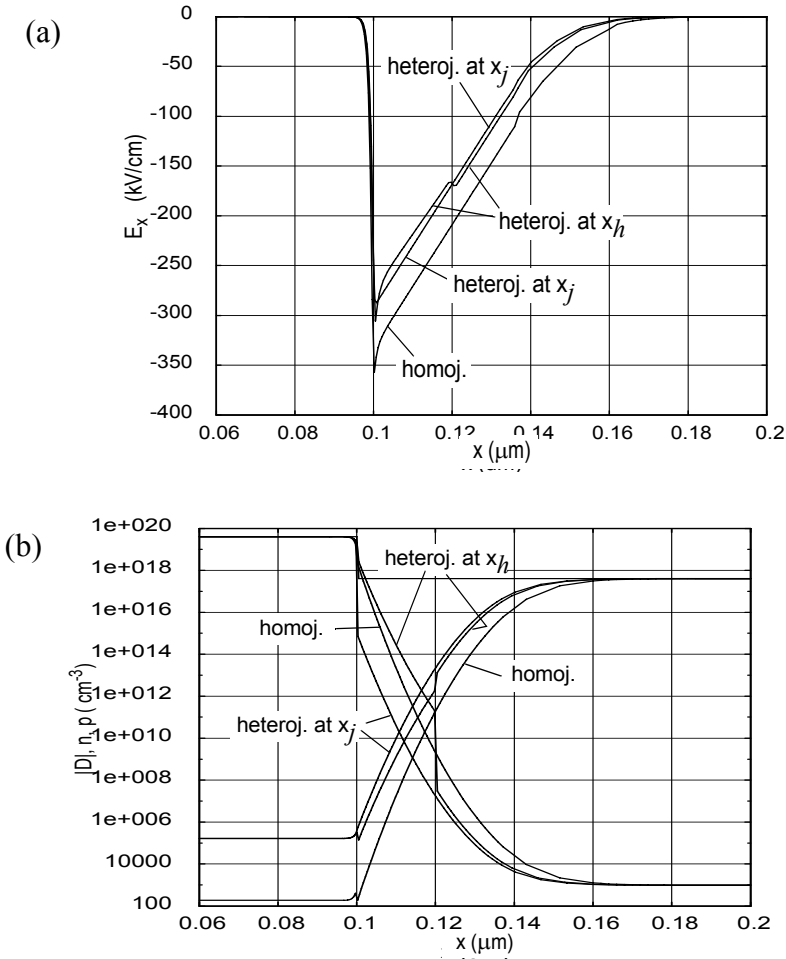


Fig. 4.5: Device simulation results in thermal equilibrium for a 1D pn diode with abrupt profiles. (a) Electric field and (b) carrier densities. Comparison between heterojunction coinciding with pn-junction  $x_j$  ( $= 0.1 \mu\text{m}$ ) heterojunction moved 20 nm into the  $n$  region and homojunction as reference. The narrow spike at  $x_j$  results from the abrupt doping profile.

#### 4.1.3 SiGe HBT

As discussed already in sec. 2.1, adding Ge to the base region of a bipolar transistor provides an additional degree of freedom to improve the elec-

trical device characteristics. Figure 4.6a shows the three basic Ge profiles encountered in practical HBTs. For the sake of simplicity, the profiles have been aligned with the junctions. Curve 1 corresponds to the classical triangular profile used in first generations of SiGe HBTs, while curve 3 is used mostly in low-emitter concentration (LEC) HBTs. Curve 2 represents a combination of the above and has become more widely used in recent HBT generations. It is also employed in the CED HBT (cf. Fig. 2.6) used throughout this book and is shown again for future reference in a magnified view in Fig. 4.6b.

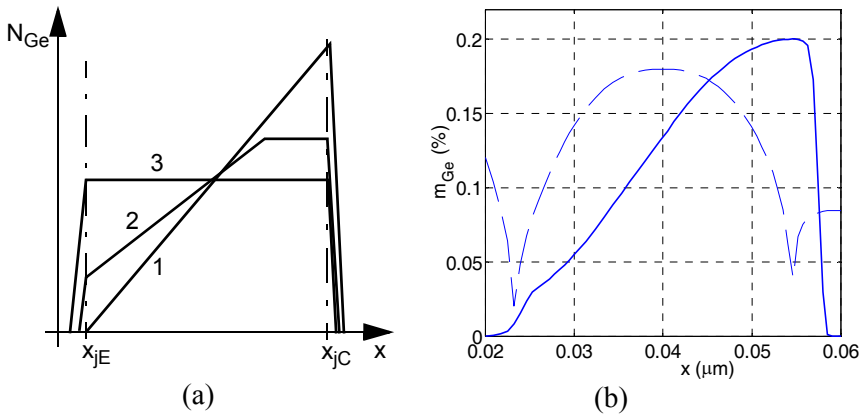


Fig. 4.6: (a) The three basic types of the Ge profile found in practical HBTs: triangular (curve 1), trapezoid (curve 2), and box (curve 3) profile with the same Ge content. (b) Magnified view of the actual Ge profile (solid line) used in device simulation for the CED HBT. The base region and junctions are indicated by the doping profile (dashed line).

The Ge profile in Fig. 4.6b was constructed with the following goals in mind:

- A sufficiently high Ge step at the BE junction prevents hole injection into the emitter, thus reducing the emitter minority charge and increasing the current gain.
- The Ge grading from the E to the C side of the base region creates a drift field that accelerates the electrons towards the collector. For a given base width, the magnitude of the drift field can in principle be adjusted by the maximum amount of Ge; roughly, a potential change of 70mV can be achieved per 10% Ge change. However, the maxi-

mum Ge contents is limited by lattice stability criteria [4,5] in order to keep the lattice from relaxing. More specifically, the overall amount of Ge determines the layer thickness over which a strained lattice can be maintained in a stable way. The advantages of a strained lattice are mobility enhancements in a certain doping range [6-9].

- From the limitation above follows that the Ge contents needs to drop to zero at some point  $x_{ch}$  in the collector. This negative slope corresponds to a retarding field for electrons. Therefore,  $x_{ch}$  has to be chosen very carefully. As described in [2,3] and evident from the discussions in section 4.1.1,  $x_{ch}$  should not occur in the neutral base region. Therefore, usually a point in the BC SCR is selected because the electrostatic field across the SCR is large enough to mask the Ge related field, at least for sufficiently low current densities.

The addition of Ge in the base region allows not only to increase the base doping but also to shift its peak away from the BE junction toward the BC junction. This has various consequences regarding carrier transport:

- On the emitter-sided base region the doping related retarding drift field decreases, but is usually more than compensated by grading the Ge contents.
- On the collector-sided region the accelerating doping related field increases, thus allowing to limit the maximum Ge contents.
- The higher doping increases the bandgap-narrowing (BGN) in the base. In principle BGN has the same effect as adding Ge. Therefore, bandgap changes associated with the doping change create an additional field, which has an accelerating effect on carriers on the emitter side and a retarding effect on the collector side. The bandgap *change* is usually larger on the collector side due to the lower collector doping compared to the emitter doping.

All of the three field components superimpose each other. Therefore, the profile in a SiGe HBT has to be very carefully designed in order to avoid undesired effects in the electrical characteristics.

The band diagram of the HBT is shown in Fig. 4.7a. The small step at the BE junction is absorbed in the potential distribution of the BE SCR. However, the impact of the Ge gradient in the base region is clearly visible in the conduction band and so is the fairly abrupt Ge decrease at the BC junction in the valence band.

The gradient in the conduction band can be explained along the same lines as in the previous sections. The band diagram in the base region with

an assumed constant doping concentration is magnified in Fig. 4.7b. For a linear Ge increase from E to C the left side shows the initial band lineup before equilibrium is restored. For negligible change in affinity, the conduction band is parallel to the macro-potential. Hence the full bandgap change initially appears in the valence band. However, since no current flow exists in equilibrium a redistribution of charge needs to bring the gradient of the Fermi-potential to zero. Therefore, majorities (holes) start flowing from the wide bandgap side at the E to the narrow bandgap side at the C, causing a negative electric field that aids electron transport. The electric field corresponds to a negative slope of the electrostatic macro-potential and, hence, the conduction band edge. This slope is also visible and indicated in Fig. 4.7a.

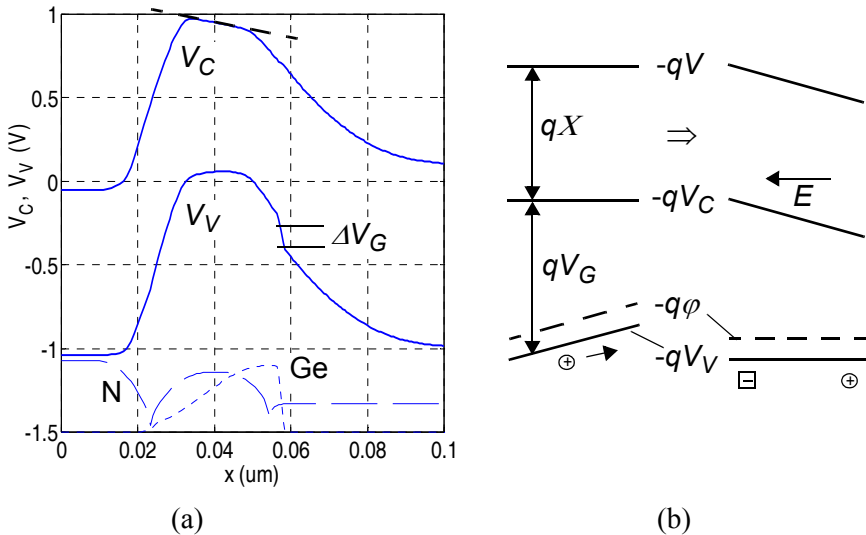


Fig. 4.7: (a) HBT band diagram at equilibrium with net doping concentration and Ge profile (not drawn to scale) as reference. (b) Formation of the equilibrium band diagram in the base region.

The current carrying capability of a SiGe HBT is eventually limited by the behavior in the collector. The effects occurring at higher current densities will be discussed further down in more detail. For understanding the limitations of SiGe HBTs it is useful to closely inspect the field and carrier density distributions in the base-collector region. Device simulation results



for the two cases of high BC reverse voltage and a forward biased BC junction are considered below. Based on these observations and the evolution of the band diagram at the BC junction with increasing current density, conclusions regarding the electrical behavior of SiGe HBTs can be drawn.

#### 4.1.3.1 High-voltage case

Figure 4.8 shows the electrostatic field and the corresponding carrier densities for selected current densities. The selection criteria are the same as for the BJT in sec. 3.2 and so is the behavior at sufficiently low current densities (curves 1 to 3). The only difference is the “discontinuity” in the electron density at  $x_{ch}$  which is clearly visible in curves 2 and 3. Like for the BJT, a horizontal field still leads to almost saturation velocity. It is interesting to note though that even the slight velocity reduction at bias point 3 causes already an increase in (base) transit time and a slight drop of  $f_T$  from its peak. Further increase of  $I_C$  causes the field slope to reverse its sign as for the BJT. For bias point 4 the extrapolated field at the junction reaches approximately  $E_{lim}$ . In contrast to the BJT though a field “spike” is visible at  $x_{ch}$ , which results from the negative Ge slope. This spike becomes clearly visible at bias point 5, with the field now assuming *positive* values within the region of the steep Ge drop. Obviously, a positive field poses a barrier for electrons entering the collector.

Another difference to the BJT is the behavior of the carrier densities at high injection (curve 4 and 5). For the HBT, there is no neutrality in the collector and the adjacent base region although the field has completely collapsed and some hole injection into the collector can be observed. The reason for this is the rapidly increasing bandgap at  $x_{ch}$ , which significantly reduces the hole density on the collector side of  $x_{ch}$ .

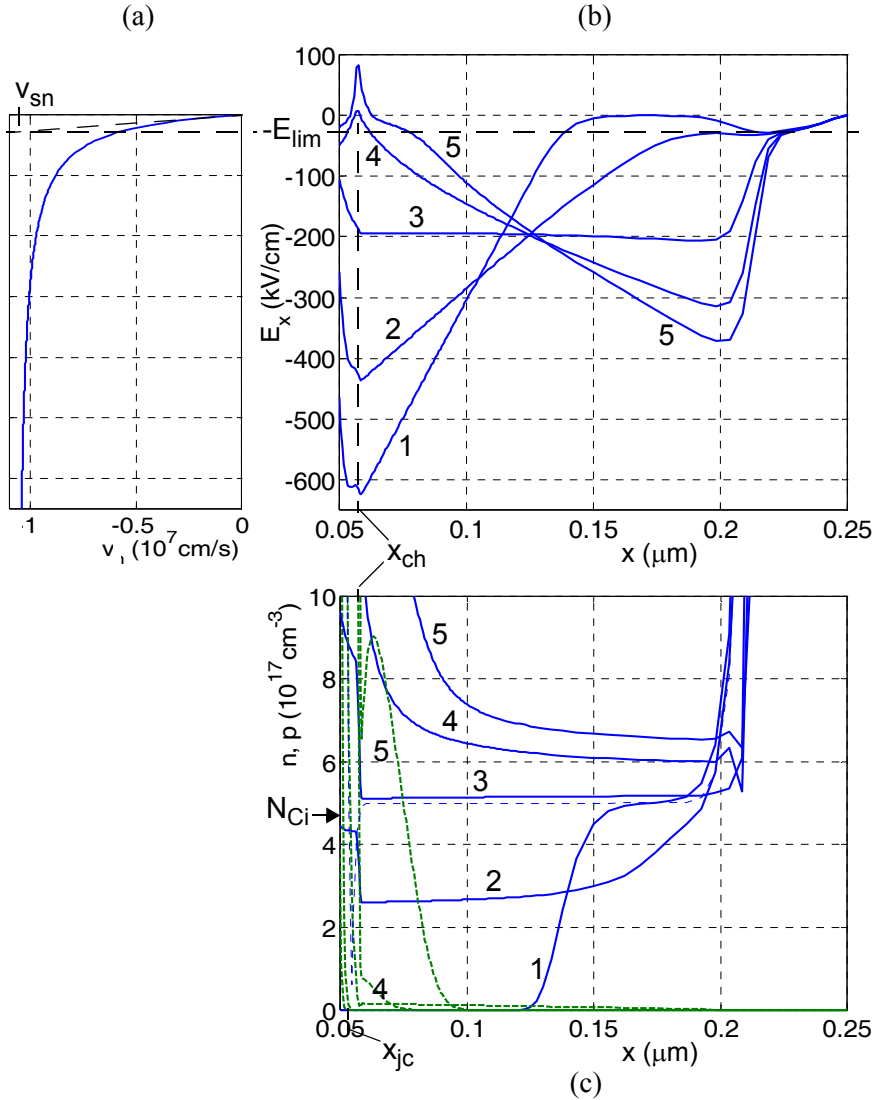


Fig. 4.8: Results of CED HBT device simulation within the base-collector region in case of a "high" voltage  $V_{ci}$ : (a) electron drift velocity  $v_n$ ; (b) electric field; (c) electron density (solid lines), hole density (dashed lines) and doping (dotted line). The curves 1...5 are for different currents  $I_C = (2.15 \cdot 10^{-3}, 4.3, 7.9, 9.7, 10.7) \text{ mA}$ ;  $V_{B'C} = -2.3 \text{ V}$  and  $E_{lim} = 28 \text{ kV/cm}$ .

#### 4.1.3.2 Low-voltage case

Figure 4.9 contains the field and carrier density distribution for selected bias points and a forward biased BC junction. At low currents (curve 1) only a small portion of the collector is depleted due to the relatively high collector doping. With increasing current (curve 2) the SCR collapses for the same reasons as in a BJT: (i) the compensation of the collector doping related space charge by mobile carriers (cf. curve 2 in Fig. 4.9c) and (ii) the voltage drop across the undepleted ohmic collector region (non-zero field). At this bias point the carrier velocity at  $x_{jC}$  has already decreased by about 10% from saturation and, hence, the transit time has increased. Therefore,  $f_T$  at bias point 2 has already dropped slightly from its peak value.

Further increase in current leads at bias point 3 to a constant field in the collector at a value that in this case is slightly lower than  $E_{lim}$ . This corresponds to an entirely ohmic collector region. Again, up to this point qualitative differences to the behavior of BJTs are hardly visible. This changes at bias point 4, where the built-in field caused by the negative slope in Ge contents at  $x_{ch}$  becomes exposed. Again, as for the high-voltage case, the positive field poses a barrier for electrons trying to enter the collector.

At bias point 4 in Fig. 4.9c an injection of holes into the collector is observed. However, due to the bandgap difference between the left- and right-hand-side of  $x_{ch}$  the hole density drops almost abruptly to values much smaller than those required for charge neutrality in the collector. The consequences of this charge imbalance will be discussed in detail in the next section.

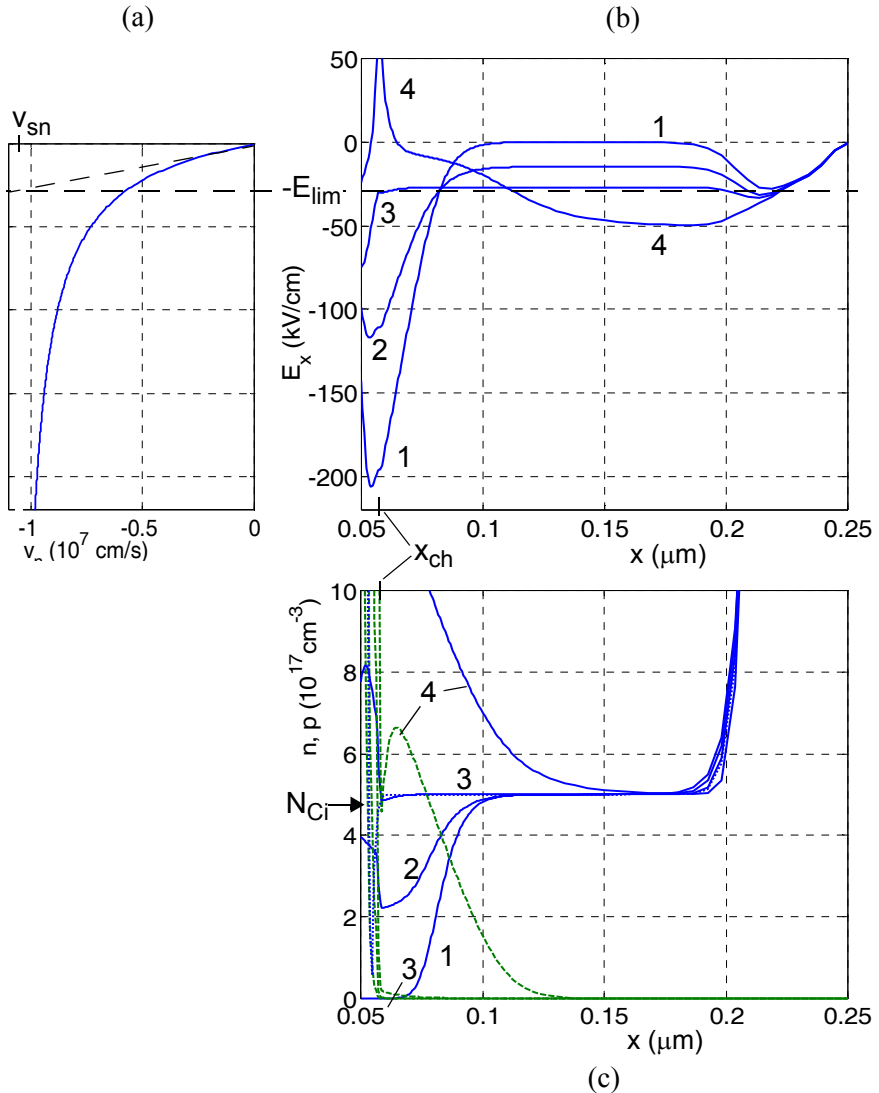


Fig. 4.9: Results of CED HBT device simulation within the base-collector region in case of a "low" voltage  $V_{Ci}$ : (a) electron drift velocity  $v_n$ ; (b) electric field; (c) electron density (solid lines), hole density (dashed lines) and doping (dotted line). The curves 1...4 are for different currents  $I_C = (2.15 \cdot 10^{-3}, 3.16, 4.55, 5.9) \text{ mA}$ ;  $V_{B'C'} = 0.4 \text{ V}$ . Note, that the residual field in the base is caused by the built-in field from doping, BGN, and material composition.

#### 4.1.3.3 The base-collector barrier

The previous discussion indicates that collector high-current effects are responsible for the deterioration of the transistor performance. This observation needs to be investigated in more detail for SiGe HBTs. The most significant difference to BJTs is the appearance of the positive (retarding) electric field caused by the Ge drop. The corresponding impact on carrier flow and charge is discussed below.

Figure 4.10 contains a magnified view of the band edges within the BC junction region for the bias points 2...5 in Fig. 4.8. According to section 4.1.1, at equilibrium the SiGe material system produces mainly a change in the valence band that poses a barrier for hole current flow. This is still valid at sufficiently low current densities: at bias point 2 the band edges rapidly drop towards the collector from their highest point in the base, with the drop in the valence band being much larger than in the conduction band. With increasing current densities, the slope of the band edges into the collector decreases. However, at bias point 3 the barrier for holes can still be clearly seen.

When further increasing the current the slope of both band edges flattens out towards  $x_{ch}$  from both sides. For the valence band the barrier remains visible but appears to shrink. For sufficiently high currents, the slope actually reverses around  $x_{ch}$  as can be seen for bias points 4 and 5. In the conduction band there is a visible step in  $V_C$  over the same region where also the Ge decays strongest. Such an increase of  $V_C$  corresponds to a barrier  $\Delta V_{Cb}$  for electrons entering the collector, which also increases with current. At the same time the barrier in the valence band shrinks, allowing more holes to be injected into the collector. The formation of the conduction band barrier is a.k.a. *collector barrier effect* in the literature (e.g. [10, 11]).

The physical reason behind the bias dependent formation of  $\Delta V_{Cb}$  can be explained by means of Fig. 4.11. At low current densities (curve 2), the negative peak of the space charge caused by the ionized doping in the base is clearly visible. This space charge disappears with increasing current and becomes positive close to the junction  $x_{jc}$  and beyond until  $x_{ch}$ . In this region the valence band barrier blocks holes from entering the collector, but there is no ionized base doping density any more to compensate these

holes, and the electron density is still too small. In fact within  $x \in [x_{jc}, x_{ch}]$  the positively charged donors in the collector add to the hole charge, leading to a thin layer with positive charge that is indicated by the hatched area left of  $x_{ch}$  in Fig. 4.11.

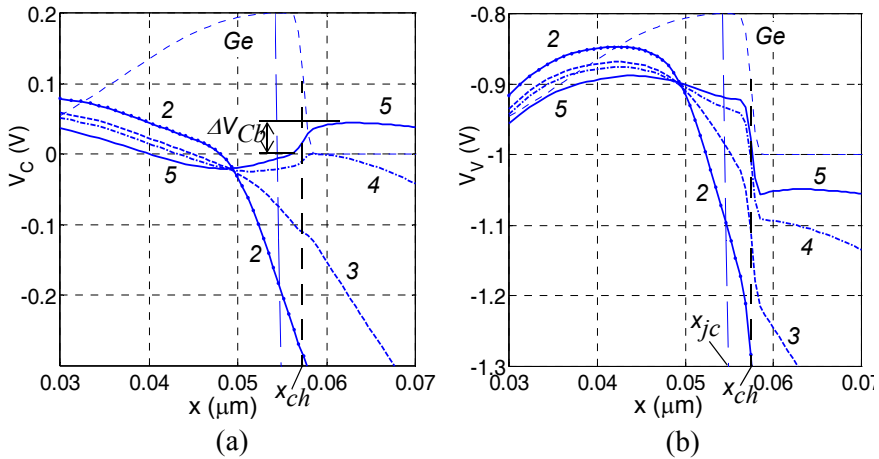


Fig. 4.10: Spatial dependence of (a) the conduction band edge and (b) valence band edge of the CED HBT for the bias points 2...5 in Fig. 4.8 and  $V_{B'C'} = -2.3\text{V}$ .

On the other side of  $x_{ch}$  the electron density reaches the collector doping at bias point 3 leading to a complete compensation of the space charge there and the extension of the SCR width to the buried layer (cf. also Fig. 4.8). With increasing current the electron density not only exceeds  $N_{Ci}$  but also has a negative gradient starting from  $x_{jc}$  to support the additionally required diffusion current. However, in contrast to BJTs, the electrons are not compensated by holes injected into the collector, leading to a thin layer with negative charge on the right of  $x_{ch}$  that is also indicated by the hatched area in Fig. 4.11.

The overall result is the formation of a so-called *dipole layer* at  $x_{ch}$ . As evident from curves 4 and 5 in Fig. 4.11, the dipole layer induces a positive electric field that corresponds to a barrier for the electron flow and causes the bending of the conduction band edge observed in Fig. 4.10. Since this field has the opposite effect on holes, it contributes to a reduction of the originally existing barrier in the valence band. As a consequence, the dipole layer charge and associated barrier  $\Delta V_{Cb}$  can increase only up to a

certain point. If at very high current densities  $\Delta V_{Cb}$  would equal the original barrier  $\Delta V_V$  in the valence band, the dipole charge would completely disappear and so the reduction of the barrier  $\Delta V_{Vb}$  in the valence band, which would be inconsistent. Note, that at such high current densities, even a slight carrier imbalance causes a large space charge density that maintains the dipole layer.

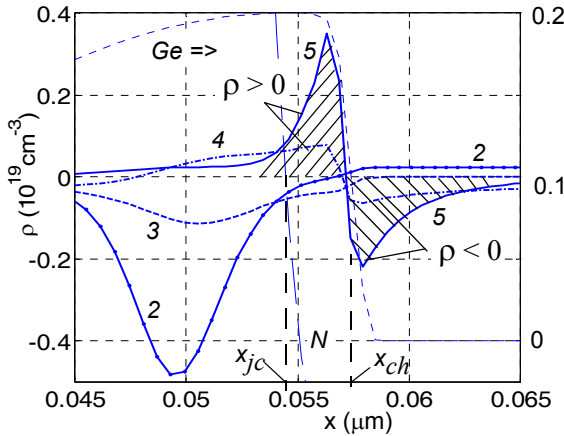


Fig. 4.11: Space charge distribution and dipole layer formation in the CED HBT for the bias points 2...5 of Fig. 4.8.

Since BGN causes the conduction band to increase already in the base region (cf. curves 4,5 in Fig. 4.10), the Ge induced barrier height is defined by the potential drop within the region *around* the steepest Ge drop as depicted for curve 5 in Fig. 4.10. Despite this somewhat arbitrary definition, the evaluation of the band edges from device simulation provides sufficient information on the formation of the barrier with bias. As Fig. 4.12 shows, the barrier increases very rapidly beyond a certain current that in turn depends on  $V_{B'C'}$ . The increase of the barrier correlates very well with the sharp increase of the minority carrier transit time, which has also been inserted in Fig. 4.12, as well as with the sharp drop of the transconductance.

According to the discussion above the barrier effect is “triggered” by high-current effects in the collector. Therefore, the relations for the electric field derived in sec. 3.2 are still expected to be applicable up to almost the critical current  $I_{CK}$ . The exact definition of the latter may have to be re-

vised due the barrier effect. In order to obtain a feeling though for its applicability,  $I_{CK}$  was calculated from the physical and technological parameters of the transistor and is also inserted in Fig. 4.12 as filled circles. Obviously,  $I_{CK}$  is a fairly good indication for the onset of the barrier effect. In fact, taking  $0.9I_{CK}$  would predict very accurately the beginning of the increase of the barrier  $V_{Cb}$ .

A very similar behavior of the barrier and the correlation to electrical characteristics is obtained for a high-voltage (power) CED HBT, which only differs by a lower doping and larger width of the collector. Such a transistor is employed for, e.g., power amplifiers.

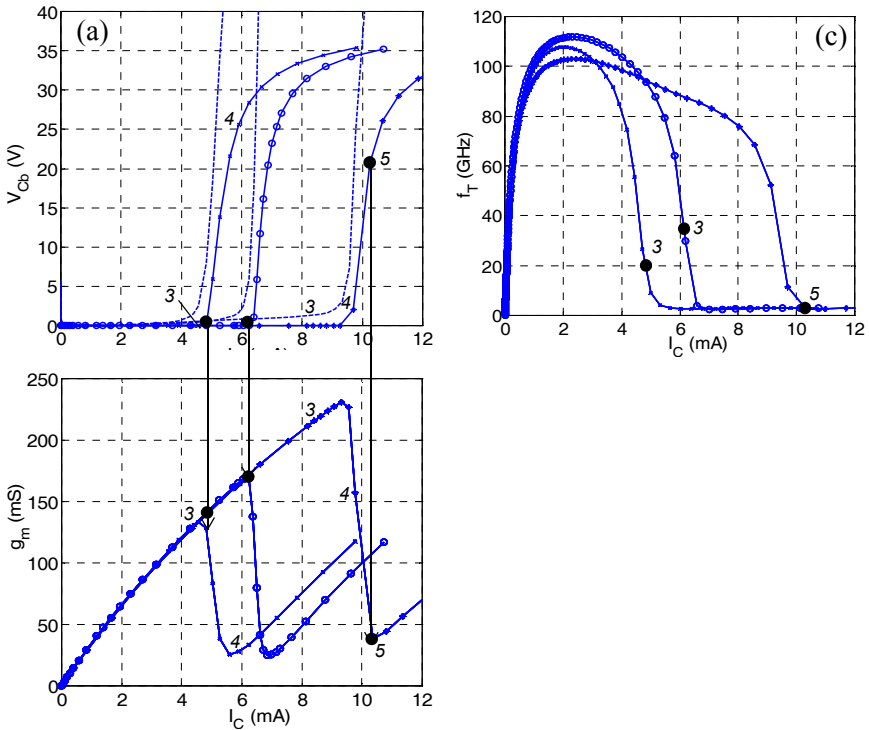


Fig. 4.12: (a) Conduction band barrier (solid lines) and minority carrier transit time (dashed lines), (b) transconductance, and (c) transit frequency vs. collector current of CED HBTs at different voltages  $V_{BE}$ : 0.4V(x), 0V(o), -2.3V(\*). The filled circles indicate the critical current  $I_{CK}$ .



The analytical treatment of the BC heterojunction has been attempted in the literature in various ways (e.g. [11,12,13]). However, in all cases an injection zone in the collector is assumed. As can be observed from the field distribution in Fig. 4.8b this situation only occurs for curve 5 (and for curve 4 in Fig. 4.9b). According to the electrical characteristics in Fig. 4.12b,c the corresponding current densities are very high and, hence, outside the practical application range. The actually interesting current range though is the one in which the barrier starts to form. Figure 4.13 shows a sketch of the development of the electric field and conduction band potential in this range with current.

For curve 1 the current density starts to cause a small spike in the field due to the dipole charge accumulation around the center (largest slope) of the Ge drop at  $x_{jc}$ , but the field  $E_{jc}$  is still negative and the extrapolated value from the slope region just reaches  $E_{lim}$ . Obviously, the small deviation from the linear behavior of the field at  $x_{jc}$  does not change the calculation for the critical current  $I_{CK}$  significantly. Once this point is reached, the field just around  $x_{jc}$  drops very rapidly with increasing current to slightly positive values first (curve 2) and then to a larger positive spike (curve 3). This corresponds to the formation of the dipole and barrier region with the width  $w_{Cb}$ . Up to this point, which is still just slightly beyond  $I_{CK}$  the barrier potential grows rapidly with current to about  $(1...2)V_T$ . Once  $\Delta V_{Cb}$  has increased and the corresponding barrier for holes,

$$\Delta V_{Vb}(I_T) = \Delta V_V - \Delta V_{Cb}(I_T)$$

has decreased sufficiently to allow holes to enter the collector, an injection region starts to form beyond the barrier region  $w_{Cb}$  (curve 4). From this point on,  $w_{Cb}$  remains more or less constant, and only the injection layer grows with current until it reaches the end of the collector at extremely high current densities. Since the exact limit of the barrier height at such high currents is of no practical interest, one can just assume  $\Delta V_{Cb}(I_T \rightarrow \infty) = \Delta V_{Vb}(I_T \rightarrow \infty) = \Delta V_V/2$ .

During the formation of the barrier region, the electric field in the collector can be assumed according to curve 2 and 3 in Fig. 4.13. Neglecting both the small area under the field within  $w_{Cb}$  and the field penetration into base and buried layer one obtains the condition

$$V_{ci} + \Delta V_{Cb} = - \int_{w_{Cb}}^{w_{Cb} + w_{Ci}} E(x) dx'. \quad (4.1)$$

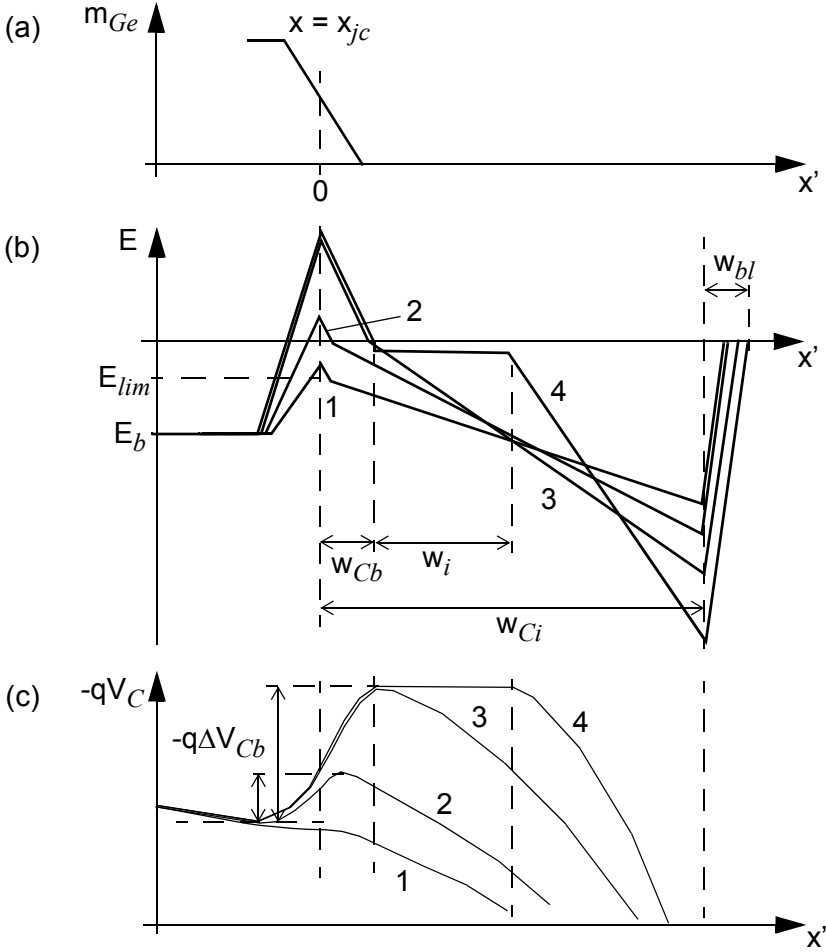


Fig. 4.13: Sketch of (a) the electric field distribution, (b) conduction band, and (c) Ge contents in the BC region for analytical treatment of the barrier potential.

However, solving Poisson's equation for the collector region  $x \in [w_{Cb}, w_{Cb} + w_{Ci}]$  does not lead to a solution that gives reasonably accurate results for  $\Delta V_{Cb}$  as function of current.

Once the injection zone has formed (curve 4 in Fig. 4.13), the assumptions of classical theory in section 3.5.3.1 can be applied to calculate the bias dependence of the barrier. In particular, space-charge neutrality is assumed within  $w_i$  allowing to write (3.267)

$$I_T = 2qA_E\mu_{nC0}V_T\frac{dn}{dx}, \quad (4.2)$$

which holds for  $2p \gg N_{Ci}$ . Inserting the linear decrease of  $n(x)$  (cf. (3.271) with  $n(x) = p(x)$ ) and setting  $n(w_{Cb}+w_i) = I_T/(qA_E v_{sn})$  yields

$$I_T = qA_E \frac{v_{sn}}{1 + \frac{w_i v_{sn}}{2\mu_{nC0}V_T}} n(w_{Cb}). \quad (4.3)$$

One difference to classical theory is that the hole density at the edge  $w_{Cb}$  of the injection zone is now calculated as

$$p(w_{Cb}) \approx N_A \exp\left(-\frac{\Delta V_V - \Delta V_{Cb}}{V_T}\right). \quad (4.4)$$

This corresponds to a situation in which the hole density has piled up in front of the barrier on the base side and has completely compensated the former SCR in a base with spatially constant base doping. The hole density on the other side of the barrier is reduced by the (now bias dependent) barrier. Inserting  $p(w_{Cb})$  from (4.4) into the neutrality condition leads to an expression for the barrier potential

$$\Delta V_{Cb} = \Delta V_V + V_T \ln\left(\frac{n(w_{Cb}) - N_{Ci}}{N_B}\right), \quad (4.5)$$

in which  $n(w_{Cb})$  from (4.3) contains the current dependence. With few rearrangements one finally obtains

$$\Delta V_{Cb} = \Delta V_V + V_T \ln\left[\left(\frac{I_T}{I_{lim}}\left(1 + \frac{V_{lim}}{V_T} \frac{w_i}{w_{Ci}}\right) - 1\right) \frac{N_{Ci}}{N_B}\right]. \quad (4.6)$$

The term in brackets is negative at low currents, then becomes positive and smaller than 1 when  $I_T$  exceeds  $I_{lim}$ . This result was also obtained in [11], just with the original physical parameters instead of  $I_{lim}$  and  $V_{lim}$ . Note, that the equation does contain an upper bound for the barrier and, hence,

does not give the correct limit at very high current densities. However, this is less relevant than the fact that the expression is only applicable *beyond*  $I_{lim}$  and, thus, can only be used for the case of high collector voltages, and *after* the barrier region has already formed.

Any solution existing in literature is based on similar assumptions as above and, hence, suffers from similar limitations. Besides being inaccurate or not applicable in the region of practical interest, the existing equations also require significant effort for smoothing and numerical conditioning. It turns out that an accurate analytical solution of  $\Delta V_{Cb}$  as function of bias is very difficult to obtain, especially in the narrow current range during barrier formation. Such a solution also depends on details of the doping and material composition at the BC heterojunction. Therefore, it seems to be more suitable for compact modeling to find a simple semi-empirical description that is based on the qualitative behavior of the barrier and contains the relevant physics-based variables. This shall be attempted below.

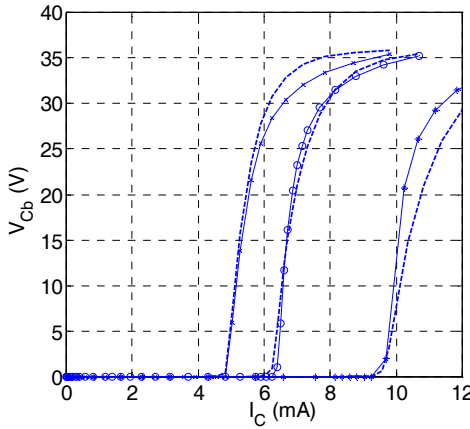


Fig. 4.14: Comparison between barrier potential from (4.8) (dashed line) and from device simulation (solid line with symbols) for the CED HBT. Parameters:  $V_{Cbar} = 0.036\text{V}$ ,  $a_{Cbar} = 0.01$ ,  $\beta_{Cbar} = 8$ .

Barrier formation is triggered by the critical current  $I_{CK}$ , which is clearly visible from Fig. 4.12. After a sharp increase the barrier must approach a saturation value for very high currents. The argument  $1 - I_{CK}/I_T$  allows such sharp increase, if an exponent factor  $\beta_{Cbar}$  is added,

$$i = 1 - \left( \frac{I_{CK}}{I_T} \right)^{\beta_{Cbar}}. \quad (4.7)$$

Then, the desired behavior is accomplished by the smoothing function

$$\Delta V_{Cb} = V_{Cbar} \frac{i + \sqrt{i^2 + a_{Cbar}}}{1 + \sqrt{1 + a_{Cbar}}}. \quad (4.8)$$

As the comparison with the barrier potential obtained directly from device simulation shows in Fig. 4.14, excellent agreement can be obtained with the simple and flexible expression above. In this case, the parameter  $V_{Cbar}$  has been kept close to  $\Delta V_I/2$ , i.e. half of the original valence band barrier. The value of  $\beta_{Cbar}$  turned out to be between 3...5, and  $a_{Cbar} \ll 1$ .

## 4.2 Electric field in the collector region

The previous observations have shown that the electrical characteristics of both BJT and HBT at medium and high current densities strongly depend on the electric field distribution in the collector. The current dependence of the field causes not only the minority carrier storage time but also the base-collector depletion capacitance and the avalanche multiplication factor to vary with current. Therefore, an efficient analytical description of the electric field in the collector region appears very attractive for physics-based compact modeling.

In sec. 3.2 simple analytical relations for the electric field  $E_{jc}$  at the BC junction were derived based on assuming either a fully depleted or a partially depleted collector. The resulting different form of the solutions cannot be used directly in a compact model. Also, they are difficult to connect by a smoothing function in order to take into account the current and voltage dependent transition from partially to fully depleted collector (and vice versa). In this section, a more general solution will be derived that has the potential to be used in compact modeling. For convenience, the 1D coordinate system starts at the BC junction with  $x' = 0$ .

The need to partition the collector into ohmic and space charge regions can be avoided if the field dependent drift velocity is used in the Poisson equation. As it turns out, the most simple form,

$$v_n(E) = v_{sn} \frac{E(x')}{E_{lim} + E(x')}, \quad (4.9)$$

represents the velocity-field dependence and trend reasonably well. Insertion into Poisson's equation yields the first-order differential equation

$$\frac{dE}{dx'} = a - \frac{b}{E} \quad (4.10)$$

with the spatially constant, but yet voltage and current dependent variables

$$a = \frac{qN_{Ci}}{\varepsilon} \left(1 - \frac{I_T}{I_{lim}}\right), \quad b = \frac{qN_{Ci}}{\varepsilon} \frac{I_T}{I_{lim}} E_{lim} \quad (4.11)$$

and  $I_{lim}$  from (3.4). Above equation is valid in the entire collector and as long as the transfer current is governed by drift; i.e.  $I_T = qnv_n(E)$ . The solution for the electric field reads [14]

$$\frac{E_{jc} - E(x')}{a} - \frac{b}{a^2} \ln\left(\frac{aE(x') - b}{aE_{jc} - b}\right) = x' \quad (4.12)$$

Note, that this solution was applied in [14] to the SCR only.

The boundary value for  $E_{jc}$  follows from integrating  $E(x')$ ,

$$V_{ci} = \int_0^{w_{Ci}} E dx' = \int_0^{w_{Ci}} E \frac{dx'}{dE} dE = \int_0^{w_c} \frac{E^2}{aE - b} dE, \quad (4.13)$$

which yields

$$V_{ci} = \frac{1}{2a}(E_{jc}^2 - E_{wc}^2) + \frac{b}{a} w_{Ci}. \quad (4.14)$$

For compact modeling one is only interested in  $E_{jc}$ . Therefore, setting  $x = w_{Ci}$  in (4.12) gives

$$a(E_{jc} - E_{wc}) - b \ln\left(\frac{aE_{wc} - b}{aE_{jc} - b}\right) = a^2 w_{Ci}. \quad (4.15)$$

For given  $I_T$  and  $V_{ci}$  equations (4.14) and (4.15) can be solved simultaneously for  $E_{jc}$  and  $E_{wc}$ .

Since a numerical solution is not very attractive for compact modeling, often an assumption is made about  $E_{wc}$  so that  $E_{jc}$  can be calculated directly from (4.14):

$$E_{jc} = \sqrt{2(aV_{ci} - bw_{Ci}) + E_{wc}^2}. \quad (4.16)$$

In order to find a suitable function for  $E_{wc}$  and to evaluate possible solutions consider Fig. 4.15, which shows the current dependence of  $E_{wc}$  and  $E_{jc}$  for different voltages  $V_{ci}$ . Two very different transistor types were chosen: a power version of the BJT with low collector doping of  $N_{Ci} = 2 \cdot 10^{16} \text{cm}^{-3}$  in Fig. 4.15a, and the CED in Fig. 4.15b. The increase of  $E_{wc}$  in the upper figures with current appears linear for the punch-through case and can be calculated from condition (3.16) and  $V_{ci} > V_{PT}$ :

$$E_{wc} = \frac{V_{ci}}{w_{Ci}} - \frac{V_{PT}}{w_{Ci}} \left(1 - \frac{I_T}{I_{lim}}\right) = \frac{V_{ci} - V_{PT}}{w_{Ci}} + \frac{V_{PT}}{w_{Ci}} \frac{I_T}{I_{lim}}. \quad (4.17)$$

The latter expression shows clearly the offset on the y-axis and the current dependence observed in Fig. 4.15a for  $V_{B'C'} = -3.3\text{V}$ . Also, it is obviously a good approximation at large voltages (and currents) for both transistor types.

At lower voltages two different current regions can be clearly distinguished. At low injection, the field in the ohmic portion of the collector is given by (3.18), which yields after inserting (4.9)

$$E_{wc} = \frac{V_{lim}}{w_{Ci}} \frac{I_T/I_{lim}}{(1 - I_T/I_{lim})} = \frac{b}{a}. \quad (4.18)$$

As Fig. 4.15 shows, this equation is a good approximation only for fairly low currents. It is valid until at sufficiently large voltages punch-through or at very low voltages a horizontal field is reached. The latter case occurs at

$$I_{CKI} = I_{lim} \frac{V_{ci}/V_{lim}}{1 + V_{ci}/V_{lim}} \quad \text{for } V_{ci} \leq V_{lim}. \quad (4.19)$$

which corresponds to (3.249) with  $\alpha_m = 1$ . At  $I_C = I_{CKI}$  corresponding points (squares in Fig. 4.15)  $E_{wc}$  from (4.18) has already increased too far. Beyond  $I_{CKI}$  high-current effects occur that make the solution (4.12) and, hence, (4.16) invalid due to the diffusion current in the injection zone. The punch-through current can be calculated from condition (4.17) by setting  $E_{wc} = E_{lim}$ ,

$$I_{PT} = I_{lim} \left( 1 - \frac{V_{ci} - V_{lim}}{V_{PT}} \right) \quad \text{for } V_{ci} > V_{lim} \quad (4.20)$$

Both  $I_{PT}$  and  $I_{CKI}$  are smaller than  $I_{lim}$ . For lower voltages the rate of increase of  $E_{wc}$  is proportional to  $V_{ci}$  since the applied voltage determines the area under the field that can be distributed and the (average) velocity drops.

Deriving an analytical description of  $E_{wc}$  that smoothly links the different cases discussed above is the task of compact model development. At this point only the fundamental approach shall be evaluated in which  $E_{wc}$  is inserted in (4.16) directly from device simulation. In order to avoid a negative argument  $u = 2(aV_{ci} - bw_{Ci}) + E_{wc}^2$  in (4.16), a simple smoothing function is used:

$$E_{jc} = \sqrt{\frac{u + \sqrt{u^2 + (k_{ejc} E_{lim})^4}}{2}}. \quad (4.21)$$

The lower portion of Fig. 4.15 shows the corresponding comparison with  $E_{jc}$  obtained from device simulation, using  $k_{ejc} = 3.3$  (power BJT) and 2.5 (HBT), respectively. At low voltages ( $V_{B'C'} = 0.4, 0V$ ) the collector is partially depleted, and the SCR moves with the square root of the voltage, leading to the root-like nonlinear decrease of  $E_{jc}$  with current that also results from (4.16) as long as  $E_{wc}$  is small. Towards higher voltages the decrease of  $E_{jc}$  becomes more linear as the SCR punches through the collector at a lower current density (cf. (4.20)). Once punch-through occurs already at  $I_T = 0$  the decrease is a quite linear, which can be clearly observed for the power BJT in Fig. 4.15a. The transition from a square-root to a linear current dependence is determined by the behavior of  $E_{wc}$ . Once in (4.16) the expression

$$u - E_{wc}^2 = 2(aV_{ci} - bw_{Ci}) = 4 \frac{V_{PT} V_{ci}}{w_{Ci}^2} - 4 \frac{V_{PT} (V_{ci} + V_{lim})}{w_{Ci}^2} \frac{I_T}{I_{lim}}$$

approaches zero, the value of  $E_{wc}$  becomes significant and is responsible for shaping the “bending” and smooth decrease of  $E_{jc}$  towards  $E_\infty$ . The comparison shows that (4.16) gives the right trend and may be suitable for



compact modeling. provided that the bias dependence of  $E_{wc}$  can be described properly.

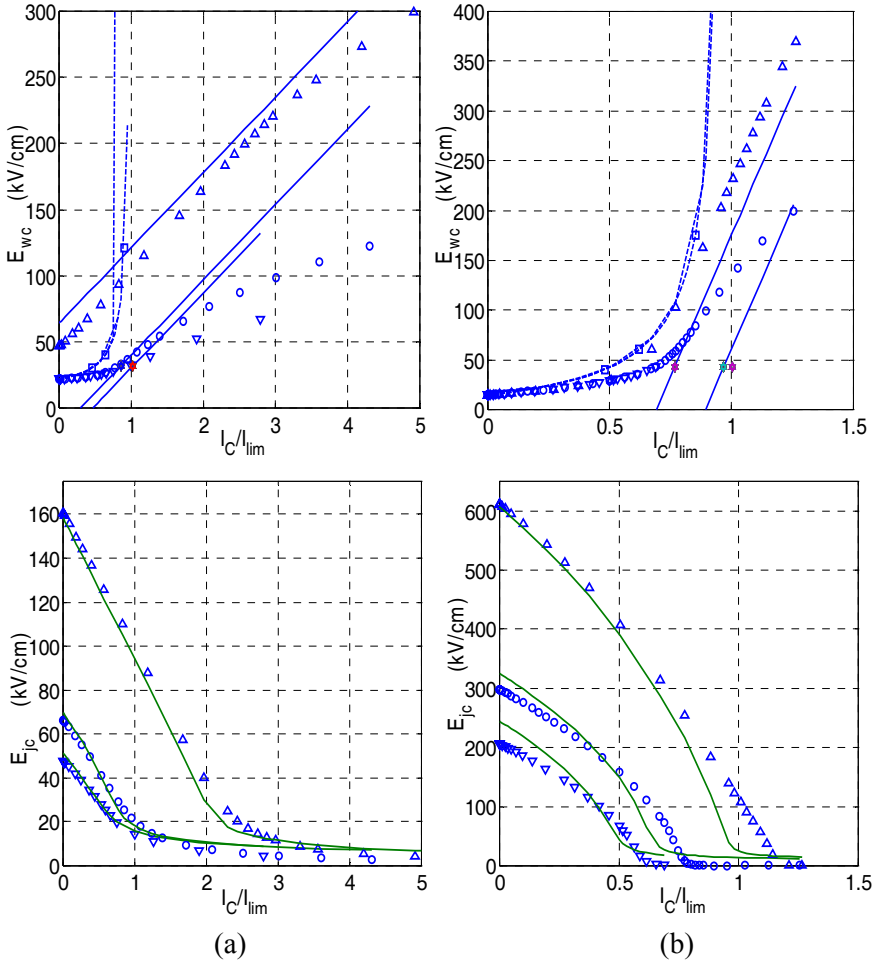


Fig. 4.15: Current dependence of  $E_{wc}$  (upper figure) and  $E_{jc}$  (lower figure) for different voltages  $V_{B'C'}$ : (a) power version of the BJT with  $N_{Ci} = 2 \cdot 10^{16} \text{ cm}^{-3}$  and for  $V_{B'C'}/V = 0.4, 0, -3.3$ ; (b) HBT with  $N_{Ci} = 5 \cdot 10^{17} \text{ cm}^{-3}$  and for  $V_{B'C'}/V = 0.4, 0, -2.3$ . Symbols represent device simulation results. In the upper figure, the solid lines represent eq. (4.17), the dashed lines (4.18), the squares (4.19), and the stars (4.20). In the lower figure the solid lines represent (4.16).

The value of  $E_{jc}$  that is reached at very high current densities can be calculated from the discussion on high-current effects in sec. 3.5.3. Assuming the Webster-effect in the injection zone, the electron transport equation (3.267) with  $N_{Ci}/2p \ll 1$  and  $dp/dx = dn/dx$  in the injection region reads

$$J_{nx}(x_{jc}) = 2q\mu_{nC0}V_T \left. \frac{dn}{dx} \right|_{x_{jc}}. \quad (4.22)$$

According to (3.269) the carrier density can be approximated quite well by a linear spatial decay, which gives with  $n(x_{jc}) \approx p(x_{jc})$  the relation

$$J_{nx}(x_{jc}) \cong -2q\mu_{nC0}V_T \frac{n(x_{jc})}{w_i}. \quad (4.23)$$

Inserting the Boltzmann relation for the electron density gives

$$J_{nx}(x_{jc}) = 2q\mu_{nC0}n_{jc} \left( \frac{d\phi_n}{dx} - \frac{d\psi}{dx} \right) \bigg|_{x_{jc}}. \quad (4.24)$$

On the other hand, the general form of the electron transport equation is

$$J_{nx}(x_{jc}) = -q\mu_{nC0}n_{jc} \left. \frac{d\phi_n}{dx} \right|_{x_{jc}}. \quad (4.25)$$

Comparison with (4.23) yields directly the gradient of the electron quasi-Fermi potential

$$\left. \frac{d\phi_n}{dx} \right|_{x_{jc}, I_T \rightarrow \infty} = \frac{2V_T}{w_i}. \quad (4.26)$$

Equating (4.24) and (4.25) allows to also calculate the gradient of the electrostatic potential

$$\left. \frac{d\psi}{dx} \right|_{x_{jc}, I_T \rightarrow \infty} = \frac{1}{2} \left. \frac{d\phi_n}{dx} \right|_{x_{jc}, I_T \rightarrow \infty} = \frac{V_T}{w_i}. \quad (4.27)$$

This result simply reflects the fact that the electron current in the injection zone is carried equally by drift and diffusion.

There are various reasons for the deviation of (4.16) from the actual value of  $E_{jc}$  observed in Fig. 4.15. First, the doping profile around the junction and buried layer is not ideal as assumed in theory, leading to a smaller

value of  $E_{jc}$ . Second, a certain fraction of the applied voltage is dropped across the base-sided portion of the BC SCR and also the transition region to the buried layer. Thus, a smaller value  $k_{vci} V_{ci}$  was already inserted in the comparisons ( $k_{vci} = 0.92$  (power BJT) and  $0.75$  (HBT)). Third, the electric field was used although in device simulation the mobility depends on the gradient of the quasi-Fermi potential. Replacing  $\psi$  by  $\varphi_n$  in (4.9) does not change the expressions obtained in this section and gives a similar trend as shown in Fig. 4.15. The values differ somewhat from  $\text{grad}(\psi)$  since  $\text{grad}(\varphi_n)$  does not contain built-in fields.

In the literature,  $E_{wc}$  is generally described by (4.18) (e.g. [14, 15]). If inserted into (4.16) this results in large deviations around and beyond  $I_{CKl}$ , which also represents the zero-crossing of the term  $(1 - I_T/I_{lim})$ . In fact, the increase of  $E_{wc}$  with current leads to a shift of the “bending” of the field from  $I_{CKl}$  to approximately  $I_{CK}$ . Employing (3.28) for  $E_{jc}$  (e.g. [15]) results in an expression that is numerically difficult to handle since using (4.20) makes it inconsistent around  $I_{CKl}$ . A simple analytical model for the bias dependence of  $E_{jc}$  was presented in [16], but is based on the punch-through expression (3.14) and, hence, leads to larger deviations for the partially depleted case.

### 4.3 Static operation and characteristics

This section deals with effects that are generally relevant in advanced bipolar transistors. The focus of the examples will be on SiGe HBTs though. The discussion starts with the derivation of integral-charge control relations with different levels of “sophistication” for the 1D DC transfer current. The final result is the 1D generalized integral-charge control relation (GICCR) with its weight factors for the various charge components. The deficiencies of classical theory and underlying reasons will be demonstrated. Finally, the dependence of the various base current components and the forward current gain on material composition and transport parameters will be discussed.

### 4.3.1 Transfer current from integral charge-control relations

Under the assumption of negligible time dependence and impact of recombination on the electron current density in the continuity equations, an accurate expression for the transfer current can be obtained for the 1D, 2D and 3D transistor structure [17, 18]. In this section, a master equation in the form of a generalized integral charge-control relation (GICCR) is derived, and possible consequences for compact modeling are discussed. The derivation is based on the 1D case to show the basic principle and to demonstrate how powerful the equation is. The discussion of a multi-dimensional GICCR is deferred to the chapter on geometry scaling theory.

Compared to the classical treatment, now the differential equation for the electron current density is solved by integration. A quick overview on the various resulting expressions can be found in [19]. The derivation starts with the electron transport equation (2.9b), which can be reformulated using the carrier density relation (2.4a) to yield the well-known form

$$J_{nx} = -q\mu_n n \frac{d\phi_n}{dx} \quad (4.28)$$

with the gradient of the electron quasi-Fermi potential as driving force. Inserting the electron density (2.4a) gives

$$J_{nx} = -q\mu_n n_i \exp\left(\frac{\psi}{V_T}\right) \exp\left(-\frac{\phi_n}{V_T}\right) \frac{d\phi_n}{dx}. \quad (4.29)$$

The effective intrinsic carrier density  $n_i$  includes bandgap changes from high-doping effects and bandgap-engineering, the impact of which will be discussed later. Furthermore, the mobility  $\mu_n$  depends on field and doping. Inserting the transformation

$$\frac{d\exp(-\phi_n/V_T)}{dx} = -\frac{1}{V_T} \exp\left(-\frac{\phi_n}{V_T}\right) \frac{d\phi_n}{dx} \quad (4.30)$$

into (4.29) leads to

$$J_{nx} = qV_T\mu_n n_i \exp\left(\frac{\psi}{V_T}\right) \frac{d\exp(-\phi_n/V_T)}{dx}. \quad (4.31)$$

Separation of differential variables and leaving only the  $\phi_n$  related term on the r.h.s. gives

$$\frac{J_{nx}}{qV_T\mu_n n_i} \exp\left(-\frac{\Psi}{V_T}\right) dx = d\left(\exp\left(-\frac{\Phi_n}{V_T}\right)\right). \quad (4.32)$$

The rather inconvenient term  $\exp(-\psi/V_T)$  can be replaced by a more suitable variable, namely the hole density  $p$ . Using (2.46) the l.h.s. is extended by the product  $n_i \exp(\phi_p/V_T)$ . This yields

$$\frac{J_{nx}}{qV_T} \frac{p}{\mu_n n_i^2 \exp(\phi_p/V_T)} dx = d\left(\exp\left(-\frac{\Phi_n}{V_T}\right)\right). \quad (4.33)$$

Integration over the general interval  $[x_l, x_u]$  gives for the r.h.s

$$\int_{\exp(-\phi_n(x_l)/V_T)}^{\exp(-\phi_n(x_u)/V_T)} d\left(\exp\left(-\frac{\Phi_n}{V_T}\right)\right) = e^{\left(-\frac{\Phi_n(x_u)}{V_T}\right)} - e^{\left(-\frac{\Phi_n(x_l)}{V_T}\right)}. \quad (4.34)$$

The exact values to be inserted for  $\phi_n$  obviously depend on the choice of the integration limits, which will be discussed later. Extension of the l.h.s of (4.33) with

$$\frac{-I_T \exp(-V_{BE}/V_T)}{A_E \mu_{nr} n_{ir}^2} \quad (4.35)$$

gives after integration

$$\int_{x_l}^{x_u} \frac{J_{nx}}{qV_T} \frac{e^{(-\phi_p/V_T)}}{\mu_n n_i^2} p dx = \frac{-I_T e^{(-V_{BE}/V_T)^{x_u}}}{A_E q V_T \mu_{nr} n_{ir}^2} \int_{x_l}^{x_u} h(x) p(x) dx, \quad (4.36)$$

with  $h(x) = h_g h_j h_e$  as the normalized weight function for the hole density and the components

$$h_g = \frac{\mu_{nr} n_{ir}^2}{\mu_n(x) n_i^2(x)}, \quad h_j = \frac{J_{nx}(x)}{-I_T/A_E}, \quad h_v = \exp\left(\frac{V_{BE} - \phi_p(x)}{V_T}\right). \quad (4.37)$$

$\mu_{nr}$  and  $n_{ir}$  are reference values taken, for instance, at some point in the neutral base region or as average value across the base at zero-bias.

Equating the results (4.34) and (4.36) yields after some rearrangement the GICCR "master" equation for the quasi-static transfer current

$$I_T = c_0 \frac{\exp\left(\frac{V_{B'E}}{V_T}\right) \left[ \exp\left(-\frac{\Phi_n(x_l)}{V_T}\right) - \exp\left(-\frac{\Phi_n(x_u)}{V_T}\right) \right]}{\int_{x_l}^{x_u} h_g h_J h_v p dx} \quad (4.38)$$

with the constant

$$c_0 = A_E q V_T \mu_{nr} n_{ir}^2. \quad (4.39)$$

Notice, that after multiplication with  $q$  the denominator integral in (4.38) has the unit of a charge density. In fact, for  $h(x) = 1$  the integral equals the hole charge density. Eq. (4.38) is the basis for the derivation of a variety of different analytical transfer current formulations. The choice of the reference values for  $\mu_{nr} n_{ir}^2$  is arbitrary and will be discussed later. Using the normalization (or reference) values  $V_{B'E}$  and  $I_T/A_E$  in (4.35) establishes in a generic way the relation to variables measurable at the device terminals. Deriving the master equation only assumes a 1D structure. So far, recombination and even a possible time dependence is included via the weight functions. A similar denominator as in (4.38) was first derived in [20]. Furthermore, bandgap differences were explicitly included in a derivation in [21], which focused just on the base region though and, hence, contained only  $h_g$ .

The validity of the master equation can be evaluated by means of device simulation. This requires a definition of the integration limits. From a pragmatic point of view, the 1D contacts are chosen here as integration limits but any other choice gives the same results for  $I_T$ . As a consequence of this choice, the electron quasi-Fermi potentials in the numerator assume their known 1D terminal values (e.g. for common-emitter configuration),

$$\Phi_n(x_l) = \Phi_n(0) = 0 \quad \text{and} \quad \Phi_n(x_u) = \Phi_n(L_x) = V_{C'E}. \quad (4.40)$$

The denominator integral is obtained directly from device simulation results and the prefactor  $c_0$  is determined at a low voltage (e.g.  $V_{B'E} = 0.7\text{V}$ ). According to Fig. 4.16, the master eq. (4.38) agrees very well with the exact results for different transistor types, namely the BJT and a CED HBT. The relative error between the transfer current calculated from (4.38) and the calculated terminal current of the simulator reaches at

maximum 0.5% at very high current densities which are far outside of the practical application range. In fact, this “error” is mostly determined by numerical integration of the denominator in (4.38). Similar results are obtained for LEC HBTs [22]. For comparison, also the results obtained from classical theory (cf. sec. 3.3) are shown, clearly demonstrating the inadequacy for modern-type transistors at medium and high current densities.

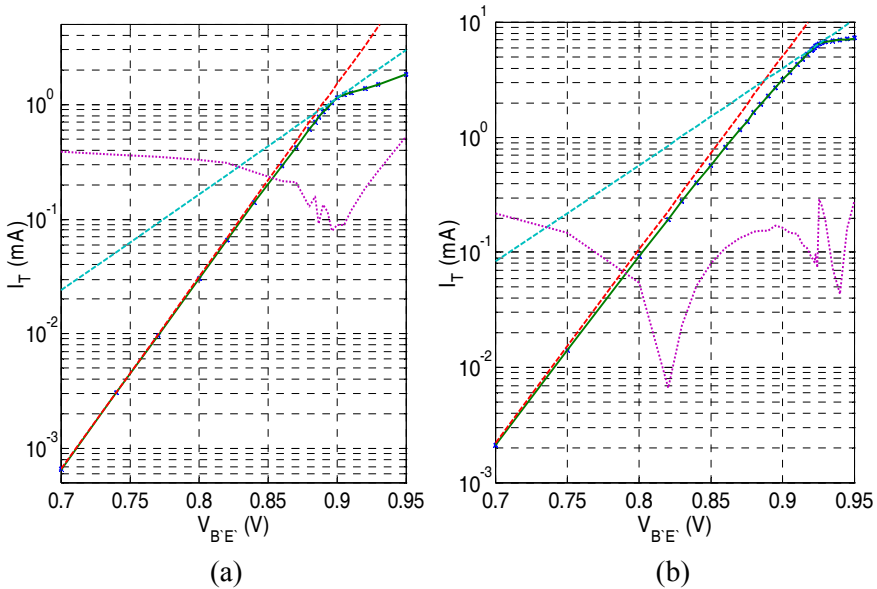


Fig. 4.16: Comparison of the analytical equation (4.38) (solid line) with the result from device simulation (symbols) and associated relative error (dotted line): (a) BJT and (b) CED HBT. The ideal characteristics at low and high injection, as obtained from classical theory or the MRR (4.42), are shown as dashed lines.  $V_{BC} = 0V$ .

Its high accuracy makes the master equation an excellent reference and a tool for a controlled development of compact transfer current relations. Generally, the choice of the integration intervals is a degree of freedom for model development. This choice influences the values of the electron quasi-Fermi potentials and the impact of the various weight functions on the integration as well as the approaches for obtaining analytical expressions. These aspects are discussed in the subsequent sections in more detail and from an application point of view.

As discussed already in sec. 3.3 classical theory leads neither to a closed-form solution for the application relevant operating region nor to a useful solution at medium and high current densities which are of particular interest for designing high-speed bipolar circuits. This difficulty is solved by the GICCR if suitable expressions for the weighted charge density in the denominator of (4.38) are inserted.

#### 4.3.1.1 The Moll-Ross relation

In classical theory, the focus was on the neutral base region, leading to  $[x_l, x_u] = [x_e, x_c]$ . Moreover, assuming constant quasi-Fermi potentials throughout the SCRs and neglecting any voltage drops toward the contacts results in

$$\varphi_n(x_e) \approx \varphi_n(0) = 0 \quad \text{and} \quad \varphi_n(x_c) \approx \varphi_n(L_x) = V_{CE},$$

so that the numerator of (4.38) reads

$$\exp\left(\frac{V_{BE}}{V_T}\right) \left[ \exp\left(-\frac{\varphi_n(x_l)}{V_T}\right) - \exp\left(-\frac{\varphi_n(x_u)}{V_T}\right) \right] \approx \exp\left(\frac{V_{BE}}{V_T}\right) \left[ 1 - \exp\left(-\frac{V_{CE}}{V_T}\right) \right]. \quad (4.41)$$

Furthermore, the hole density  $p$  is replaced by the base doping  $N_B$ , and the physical parameters are assumed to be spatially constant with  $\mu_{nB} n_{iB}^2$  as their bias independent spatial average value. Finally, neglecting recombination and time dependence, which gives  $h_J = 1$  according to (2.8b), and setting  $\varphi_p(x) = V_{BE}$ , which leads to  $h_v = 1$ , results into the Moll-Ross relation (MRR) [23]

$$I_T = A_E q V_T \mu_{nB} n_{iB}^2 \frac{\exp(V_{BE}/V_T) - \exp(V_{BC}/V_T)}{\int_{\bar{x}_e}^{\bar{x}_c} N_B dx} \quad (4.42)$$

with  $\bar{x}_e$  and  $\bar{x}_c$  indicating average values (w.r.t bias) for the boundaries of the SCRs. This relation is more general compared to the conventional approach of sec. 3.3.1 or classical transistor theory along the lines of, e.g., Ebers and Moll [24]. The above expression led to the valuable insight that the transfer current is linked to the *integral* of the base doping rather than to local values such as  $N_B(x_e)$  or  $N_B(x_c)$ .



The characteristic (4.42) is shown in Fig. 4.16, with the saturation current

$$I_S = A_E q V_T \overline{\mu_{nB} n_{iB}^2} / \int_{\bar{x}_e}^{\bar{x}_c} N_B dx \quad (4.43)$$

determined at low injection. Because of the missing bias dependence in the denominator integral, an ideal characteristic is obtained that can deviate significantly from the actual curve at medium and high forward bias. This also has significant impact on the transconductance. Except for the assumptions  $h_J = 1$  and  $h_{ve} = 1$ , which are both reasonable, all other assumptions made in the derivation of (4.42) are too crude. Hence, the MRR is not suitable for circuit design.

#### 4.3.1.2 The integral charge-control relation

Dropping the assumption of  $p = N_B$  and of the average values for the SCR edges requires a revised definition of the integration limits. Since the latter are bias dependent, a safe choice is to select the outside edges of the BE and BC SCR [25] where the hole density is negligible compared to its values in the base region. The integration limits then are  $[x_l, x_u] = [w_E, x_{Cc}]$ . The advantage of selecting these limits is also that the condition of constant electron quasi-Fermi potentials throughout the SCRs, i.e.

$$\varphi_n(w_E) \cong \varphi_n(0) = 0 \quad \text{and} \quad \varphi_n(x_{Cc}) \cong \varphi_n(L_x) = V_{CE}, \quad (4.44)$$

is much better fulfilled than in the Moll-Ross relation, at least at lower current densities.

The integral in the denominator of (4.38) is more difficult to oversee since  $h$  and  $p$  depend on both location and bias. Thus, it is instructional first to investigate the weight function and its components in relation to the hole density distribution. These results are shown in Fig. 4.17 and Fig. 4.18 for selected bias at points low injection and at high injection beyond peak  $f_T$ . The BJT and the CED HBT were chosen to visualize also the difference within the different transistor types. The weighted hole density  $h^*p$  shows the impact of the weight function on the actual hole density  $p$ . The normalization of these variables is arbitrary since just the shape and its variation

with bias is of importance. From Fig. 4.17 and Fig. 4.18 the following observations are noted:

- The negligible impact of recombination is clearly visible in  $h_J$  which - due to (2.8b) - equals 1 everywhere. The slight increase in the neutral emitter due to backinjection of holes and corresponding recombination is invisible. The largest deviation is at the contact and can reach  $B_f^{-1}$  for wide emitters. However, advanced transistors have almost transparent emitters and infinite contact recombination velocity.
- The hole quasi-Fermi potential equals  $V_{B'E'}$  over the entire base and also the adjacent space-charge regions. Hence, for all bias conditions  $h_v$  equals 1 in all regions where  $p(x)$  is significant.
- Due to  $h_J = h_v = 1$ , the component  $h_g$  determines almost entirely the weight function  $h$  for the 1D case. The effective intrinsic carrier density  $n_i$  assumes its low-doping value in the epi collector, but increases in the base and emitter region (as well as in the buried layer) due to high doping effects (in both BJT and HBT) and the composition dependence in the base region of the HBT. The electron mobility shows the opposite trend due to its doping dependence, leading to a certain compensation of the  $n_i^2$  variation. However, the latter has the dominant impact on the shape of  $h$ . The mobility also depends on the electric field which causes the weight function  $h_g$  to change with bias. Notice, that this bias dependence appears mostly in the collector where the hole density is small up to high current densities. The increase of  $h \sim 1/n_i^2$  in Fig. 4.17a across the base region toward the collector is caused mostly by bandgap narrowing, while the subsequent decrease of  $h$  at  $x \approx 0.11\mu\text{m}$  is due to the field dependence of the mobility at the end of the BC SCR. For the HBT, the sharp increase of  $h$  in Fig. 4.18a at  $x \approx 0.05\mu\text{m}$  is caused mostly by the composition dependence of  $n_i$ , while the decrease of  $h$  results again from the field dependent mobility at the end of the SCR. At high current densities, the field in this region has decreased and the mobility increased, almost eliminating this spatial dependence.
- At high injection the hole density starts to spread towards the emitter and collector region, and also increases significantly in those regions. A change in hole distribution with bias means that a spatial average value  $\bar{h}$ , determined at a particular bias point, can become bias dependent even if  $h$  does not depend on bias.

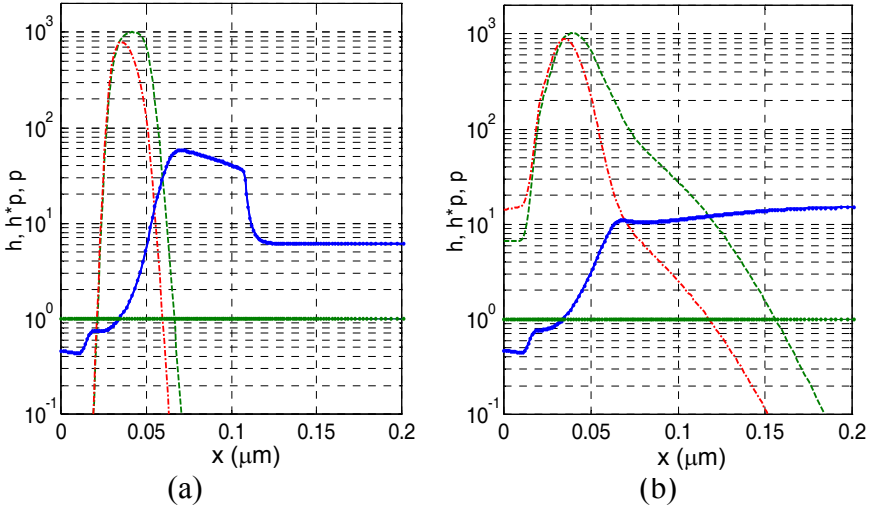


Fig. 4.17: Spatial dependence of the weighting function  $h$  (solid line), its component  $h_g$  (dots on top of  $h$ ),  $h_J$  (solid line at  $10^0$ ), and  $h_v$  (dots at  $10^0$ ) as well as the normalized weighted hole distribution  $h^*p$  (dashed line) in relation to the normalized actual hole distribution  $p$  (dash-dotted line) of the BJT at different bias conditions: (a)  $(V_{B'E}/V, I_C/\mu A) = (0.7, 0.66)$ ; (b)  $(V_{B'E}/V, I_C/\text{mA}) = (0.9, 1.15)$ .  $V_{B'C} = 0V$ . The normalization of  $h^*p$  and  $p$  is arbitrary to fit into the figure.

- The effect of the weight function on  $p$  and, hence, on the spatial distribution of the integrand  $h^*p$  in (4.38) is such that the impact of  $p$  is suppressed for  $h < \bar{h}$  and increased for  $h > \bar{h}$ . For the BJT in Fig. 4.17  $h^*p$  is shifted towards the collector with increasing bias due to high-current effects and the associated hole injection. For the HBT in Fig. 4.18 the weighted hole density is shifted w.r.t.  $p$  towards the emitter region. This shift increases with bias, indicating that variations of  $p$  in the BE region are more important than those in the BC region.

As a result of the above observations, the spatial and bias dependence of the overall weight function  $h$  is dominated by  $h_g$ , while an average value very close to 1 can be assumed for the functions  $h_J$  and  $h_v$  without introducing any significant error. This gives for the weighted total hole charge

$$Q_{ph} = qA_E \int_{x_l}^{x_u} h p dx \cong \bar{h}_J \bar{h}_v Q_{phg} = \bar{h}_J \bar{h}_v \left( qA_E \int_{x_l}^{x_u} h_g p dx \right). \quad (4.45)$$

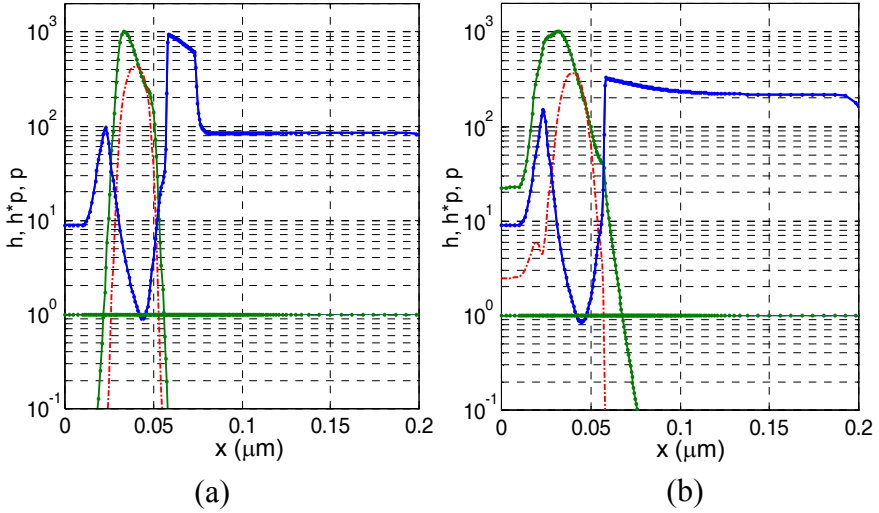


Fig. 4.18: Spatial dependence of the weight function  $h$  (solid line), its component  $h_g$  (dots on top of  $h$ ),  $h_J$  (solid line at 1), and  $h_v$  (dots at 1) as well as the normalized weight-hole distribution  $h^*p$  (dashed line) in relation to the normalized actual hole distribution  $p$  (dash-dotted line) of the CED HBT at different bias conditions: (a)  $(V_{B'E'}/V, I_C/\mu\text{A}) = (0.7, 2.1)$ ; (b)  $(V_{B'E'}/V, I_C/\text{mA}) = (0.92, 5.72)$ .  $V_{B'C'} = 0\text{V}$ . The normalization of  $h^*p$  and  $p$  is arbitrary.

Assuming an average value  $\bar{h}_g$  within the integration limits defined by (4.44) and inserting these limits also in (4.45) leads to the hole charge

$$Q_{pB} = qA_E \int_{w_E}^{x_{Cc}} p dx, \quad (4.46)$$

which represents the base region and its adjacent space charge regions. Inserting this charge together with (4.44) and  $\bar{h}_g$  into (4.38), the integral charge-control relation (ICCR) is obtained as given originally by Gummel in [26, 25],

$$I_T = c_{B0} \frac{\exp\left(\frac{V_{B'E'}}{V_T}\right) - \exp\left(\frac{V_{B'C'}}{V_T}\right)}{Q_{pB}}, \quad (4.47)$$

with the constant (and model parameter)

$$c_{B0} = qA_E \frac{c_0}{\bar{h}_j \bar{h}_v \bar{h}_g} \quad (4.48)$$

For the first time, the bias dependence of the denominator integral was captured by a bias dependent physics-based variable, namely the base hole charge.

The relation to classical transistor theory can be easily obtained by converting the prefactor  $c_{B0}$  into a saturation current

$$I_S = c_{B0} / Q_{p0}. \quad (4.49a)$$

so that

$$I_T = I_{Tf} - I_{Tr} = \frac{I_S}{Q_{pB} / Q_{p0}} \left[ \exp\left(\frac{V_{B'E}}{V_T}\right) - \exp\left(\frac{V_{B'C}}{V_T}\right) \right]. \quad (4.49b)$$

As a result, the ICCR provides - in contrast to classical theory - a closed-form approximation for the transfer current over the total operating region. According to [25] and also the implementation in the SGPM, the base charge is defined as

$$Q_{pB} = Q_{p0} + Q_{jEi} + Q_{jCi} + \tau_{Bf} I_{Tf} + \tau_{Br} I_{Tr} \quad (4.50)$$

with  $Q_{p0}$  as zero-bias hole charge,  $Q_{jEi}$  and  $Q_{jCi}$  as the 1D BE and BC depletion charge, respectively. The last two components represent the minority charges that, in the SGPM, are modeled by their low-injection formulation. Inserting  $I_{Tf}$  and  $I_{Tr}$  from (4.49b) yields a quadratic equation for  $Q_{pB}$  that can be solved explicitly.

Figure 4.19 shows a comparison of (4.47) with device simulation results for a BJT and a CED HBT. The constant  $c_{B0}$  was determined at low injection. Furthermore, the charge  $Q_{pB}$  was taken directly from device simulation in order to provide a feeling for the maximum achievable accuracy and since any analytical description is compact model specific. The results show that the ICCR is reasonably accurate for BJTs up to the onset of high current effects (at about  $I_{CK}$ ), but fails already at medium current densities for CED HBTs where especially the resulting error in the transconductance is a serious issue for practical applications. The error becomes worse for LEC HBTs and is caused mainly by neglecting the impact of the charge portions outside of the base or, equivalently, by using the 1D terminal volt-

ages - especially  $V_{B'C'}$ , instead of the quasi-Fermi potential difference  $(\varphi_p(x_{Cc}) - \varphi_n(x_{Cc}))$ .

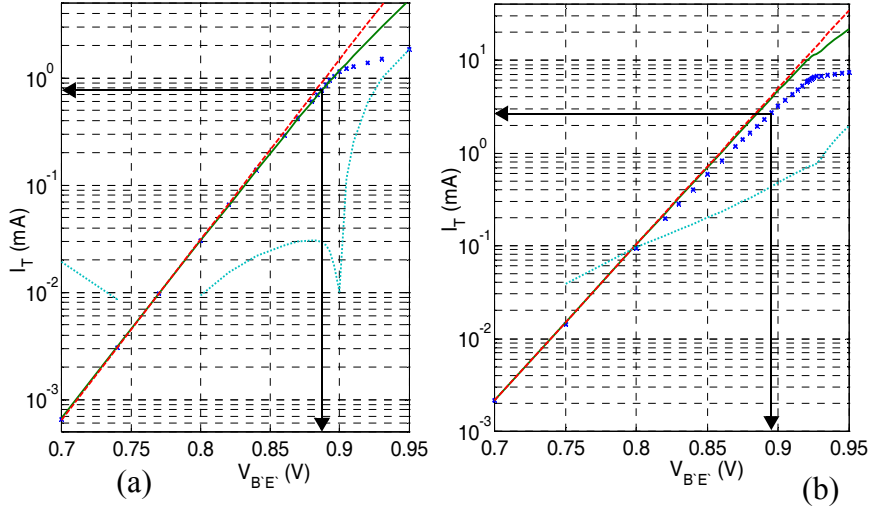


Fig. 4.19: Comparison of the transfer current obtained from the ICCR (4.47) with (4.48) (solid line) and from device simulation (symbols): (a) BJT and (b) CED HBT.  $V_{B'C'} = 0$  V. The dotted line indicates the error, the dashed line the ideal characteristic and the arrows the peak of  $f_T$ .

As a consequence of the results, it can be concluded that the ICCR is not suitable for HBT compact modeling. This statement is even more valid if analytical approximations are used for  $Q_{pB}$ , as it is the case in compact transistor models such as the SGPM. The latter would lead at high-current operation to a dependence in the form of  $\exp(V_{BE}/(2V_T))$  which is the same as in classical theory. Notice that in practice the inherent errors of the ICCR at medium current densities are often compensated by tweaking model parameters (to non-physical values). For instance, increasing the emitter resistance can reduce the deviation in the transfer characteristic and transconductance, but only at the expense of errors in the dynamic behavior.

### 4.3.1.3 Generalized integral charge-control relation

As shown in Fig. 4.16, the master equation (4.38) yields excellent agreement with the exact solution. For the comparison, the 1D contacts at  $x = [0, L_x]$  were chosen as integration limits. Major reasons for this choice are:

- The 1D terminal voltages are clearly defined and can be measured in contrast to internal differences in quasi-Fermi potentials.
- Developing suitable physics-based analytical relations for the internal potentials as function of applied bias, especially at higher current densities, has historically turned out to be very difficult.

The selected integration limits lead to (4.40) and, hence, to

$$\exp\left(\frac{V_{BE}}{V_T}\right) \left[ \exp\left(-\frac{\varphi_n(0)}{V_T}\right) - \exp\left(-\frac{\varphi_n(L_x)}{V_T}\right) \right] = \exp\left(\frac{V_{BE}}{V_T}\right) \left[ 1 - \exp\left(-\frac{V_{CE}}{V_T}\right) \right], \quad (4.51)$$

instead of the approximation (4.41). Inserting this into (4.38) gives the relation

$$I_T = qA_E c_0 \frac{\exp\left(\frac{V_{BE}}{V_T}\right) - \exp\left(\frac{V_{BC}}{V_T}\right)}{Q_{ph}}, \quad (4.52)$$

which includes the ICCR as subset.

The weighted hole charge  $Q_{ph}$ , as defined in (4.45), is not directly suitable for compact modeling and, hence, needs further investigation. An important point for practical applications is whether a variable appearing in a model equation can be experimentally determined. This can be achieved here, for instance, if the integral term can somehow be transformed into a measurable charge as in the ICCR. For this, the hole density

$$p(x) = p_0(x) + \Delta p(x), \quad (4.53)$$

is partitioned into its equilibrium component  $p_0$  and its *change*  $\Delta p$  with respect to equilibrium. Thus  $\Delta p$  contains all bias dependent depletion and minority carrier contributions. While the latter are distributed over the total transistor, carriers related to (dis-)charging the depletion regions are located around the junctions and are related to the (ionized) base doping. With the above partitioning,  $Q_{ph}$  can then be decomposed into an equilibrium related component

$$Q_{p0h} = qA_E \int_0^{L_x} h p_0 dx \quad (4.54)$$

and a bias dependent component

$$\Delta Q_{ph} = qA_E \int_0^{L_x} h \Delta p dx . \quad (4.55)$$

For a given bias point, a spatial average for the weight factor can be defined,

$$\bar{h}_0 = Q_{p0h} / Q_{p0} , \quad (4.56)$$

that relates  $Q_{p0h}$  to the actual zero-bias hole charge

$$Q_{p0} = qA_E \int_0^{L_x} p_0 dx . \quad (4.57)$$

If the impact of the field dependent mobility is negligible,  $\bar{h}_0$  would be bias independent. The normalized product  $h p_0$  looks similar to  $h p$  displayed in Fig. 4.17a and Fig. 4.18a. For medium to high injection, the portion of  $p_0$  towards the junctions is weighted differently than for zero-bias.

The bias dependence of  $Q_{p0h}$  calculated from device simulation is shown in Fig. 4.20. Both the BJT and HBT show the same trend. First,  $Q_{p0h}$  is significantly larger than  $Q_{p0}$  since  $h(x)$  is mostly larger than 1 over the base region. Second, at low injection  $Q_{p0h}$  increases with  $V_{B'C'}$  due to the steeper slope and, hence, larger values of the weight function towards the collector end of the base (cf. Fig. 4.18). This increase of  $h(x)$  is caused by the mobility reduction with field. For a given voltage  $V_{B'C'}$  the weighted charge stays constant until the transfer current starts to influence the electric field distribution at the BC junction. Third, once at high injection the field starts to drop, the mobility increases and  $h(x)$  in that region decreases, leading to a decrease of  $Q_{p0h}$ . This decrease levels off at very high current densities when the mobility assumes its low-field value. While the voltage variation of  $Q_{p0h}$  is stronger in the BJT, its current related change is overall less pronounced than for the HBT, in which the barrier effect at the BC junction plays an important role. As can be observed, the bias dependence of  $Q_{p0h}$  (and thus  $\bar{h}_0$ ) in advanced transistors can be quite sig-



nificant, so that the assumption of a constant value is justified only up to medium current densities (i.e. below peak  $f_T$ ).

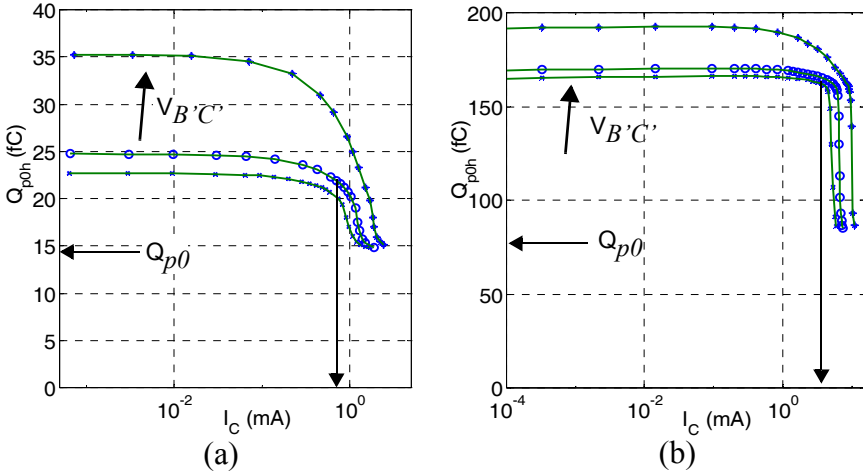


Fig. 4.20: Current dependence of weighted zero-bias charge  $Q_{p0h}$ : (a) BJT and (b) CED HBT.  $V_{B'C'}/V = 0.4$  (+), 0 (o), -2.3 (\*). The arrows indicate peak  $f_T$  at  $V_{B'C'} = 0$  V and  $Q_{p0}$ .

Compared to  $Q_{p0h}$ , the component  $\Delta Q_{ph}$  is determined by the spatial and bias dependence of both  $h$  and of  $\Delta p$ . At low injection, when the carriers are still confined to the base region, a strong spatial dependence of  $h$  results in a different weight factor for the carriers charging the depletion regions compared to the equilibrium case. Towards high current densities the shape of  $\Delta p$  can start to change significantly due to the influence of the minority carriers. Figure 4.21 shows for a BJT the distribution  $\Delta p$  and  $h\Delta p$  for different bias points. The  $BC$  voltage dependence of the respective curve shapes is displayed in Fig. 4.21a along with the net doping profile as reference. Since the BE junction is forward biased at the same  $V_{B'E'}$  for all  $V_{B'C'}$ , the same positive excess hole density appears there and compensates the original BE space charge. The weighted density  $h\Delta p$  is slightly smaller than  $\Delta p$  since  $h < 1$  in this region (cf. Fig. 4.17). In the BC region one can observe a positive excess hole density for  $V_{B'C'} > 0$ , a negligible density for  $V_{B'C'} = 0$ , and a negative density for  $V_{B'C'} < 0$ . In all cases though, the magnitude of  $h\Delta p$  is larger than that of  $\Delta p$  due to  $h > 1$  in this region (cf.

Fig. 4.17). This has obviously consequences for the weight factors related to the depletion charges.

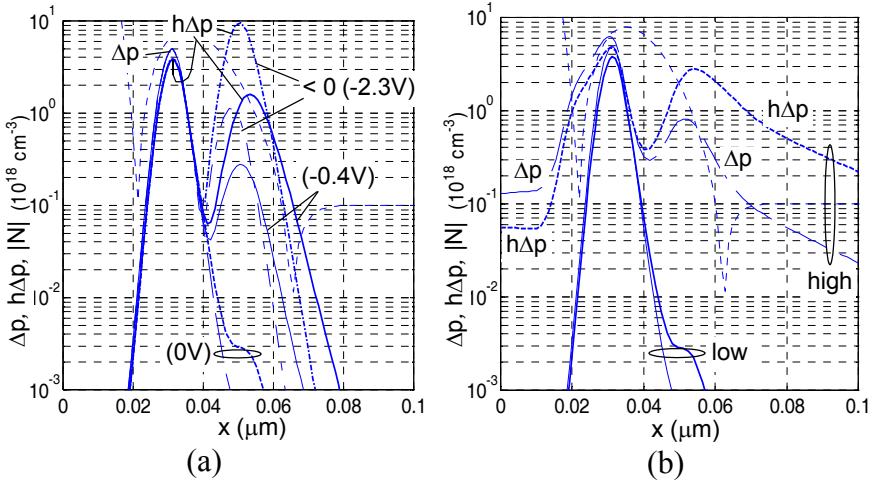


Fig. 4.21: Spatial distribution of actual hole density change  $\Delta p$  (thin lines) and its weighted counterpart  $h\Delta p$  (thick lines) for the BJT: (a) Voltage dependence at low current densities ( $V_{BE} = 0.7\text{V}$ ) for  $V_{BC} = 0.4\text{V}$  (solid),  $0\text{V}$  (dashed),  $-2.3\text{V}$  (dash-dotted). (b) Current dependence at  $V_{BC} = 0\text{V}$  for the same low and high injection bias point as in Fig. 4.17:  $I_C = 0.66\mu\text{A}$  (solid),  $1.15\text{mA}$  (dashed). As reference, the net doping profile has been inserted as dotted line.

The *current* dependent shape of excess densities is shown in Fig. 4.21b. The curves at low injection are the same as in Fig. 4.21a and serve as reference. At high injection the hole density increases significantly in all regions, leading to a different shape of the distribution. As for low injection, the hole density left from the peak, i.e. towards the emitter, is weighted with  $h < 0$ , while the hole density towards the collector region is weighted with  $h > 0$ . Hence, changes in the collector region have a larger impact on the transfer current characteristics.

Figure 4.22 contains the distribution  $\Delta p$  and  $h\Delta p$  for the CED HBT. Again, the voltage dependence in (a) exhibits two distinct regions, one at the BE junction and one at the BC junction, which correspond to the neutralized space-charge there. In contrast to the BJT, the weight factor in the BE SCR of the HBT is now significantly larger than 1, so that  $h\Delta p > \Delta p$ , as it is also the case in the BC SCR (like in the BJT). However, the weight

factor varies strongly with the Ge contents in the BE region (e.g. triangular or box Ge profile).

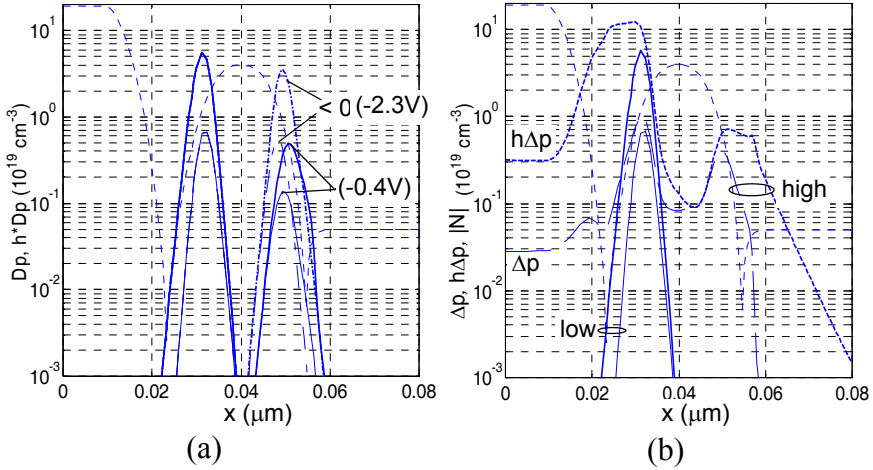


Fig. 4.22: Spatial distribution of actual hole density change  $\Delta p$  (thin lines) and its weighted counterpart  $h\Delta p$  (thick lines) for the CED HBT: (a) Voltage dependence at low current densities ( $V_{BE} = 0.7\text{V}$ ) for  $V_{BC} = 0.4\text{V}$  (solid),  $0\text{V}$  (dashed),  $-2.3\text{V}$  (dash-dotted). (b) Current dependence at  $V_{BC} = 0\text{V}$  for the same low and high injection bias point as in Fig. 4.18:  $I_C = 2.1/\mu\text{A}$  (solid),  $5.72\text{mA}$  (dashed). As reference, the net doping profile has been inserted as dotted line.

The current dependent shape of excess densities in Fig. 4.22b exhibits a spread and magnitude change which is stronger, especially in the BE region, than in the BJT for a very similar current ratio between the two cases considered. Also, in contrast to the BJT, now  $h\Delta p \gg \Delta p$  in the BE and E region. Thus a charge change in these regions has a much larger impact on the transfer current characteristic of an HBT than changes in the collector region.

It is obvious from the above results that for advanced transistors with significant BGN and change in material composition a single average weight factor for the total charge

$$\bar{h}_\Delta = \Delta Q_{ph} / \Delta Q_p \quad (4.58)$$

is strongly bias dependent and not suitable. This bias dependence is determined by the bias dependent charge contributions in the physically different transistor regions. Since  $h$  depends on location, the changing spatial

distribution of the carrier densities with bias causes the product  $h\Delta p$  to contribute to the integration in a non-proportional way; the bias dependence of  $h$  itself plays also a role which is relatively small though. Fundamentally, the partitioning of the transistor into regions and, hence, of  $\Delta Q_{ph}$  into respective components is arbitrary. The most simple way is to select metallurgical regions while a more sophisticated partitioning may follow the Regional Approach. Also, one could start with a small-signal description and then obtain  $\Delta Q_{ph}$  from an integration of weighted small-signal quantities. Moreover, rather than trying to use averaged weight factors as in (4.58) and to model  $\Delta p$  separately, one can try to directly describe the weighted carrier density

$$\Delta p_h = h\Delta p \quad (4.59)$$

as a function of bias, temperature etc. (e.g. [27]).

Whatever approach is chosen though needs to provide an *analytical* description of *measurable* variables. This practical requirement narrows down the possible options. A viable and so far successful approach is discussed below. The goal is to reduce the weighted charge components to a product of the actual (measurable) charge components and corresponding spatial averages of weight factors so as to use the already existing bias dependent description of the actual charges, which still determine the fundamental dependence on bias, temperature and process parameters. Since at some point simplifications are necessary to arrive at *compact* analytical equations, exact transformations from  $\Delta Q_{ph}$  expressions to expressions containing exact definitions (according to, e.g., the Regional Approach) of small-signal quantities are sought. This way, the problem is subdivided into two steps:

- Establish a relation between weighted charge components and their corresponding actual counterparts, which can both be calculated from device simulation results. This allows to calculate the exact weight factors and, hence, to evaluate the error made by approximating them with simplified equations (such as constant values).
- Replace the actual charges by compact analytical equations that include small-signal variables, such as capacitances and storage times, which can be measured via the terminals. Their derivation can be performed separately based on, e.g., a Regional Approach that can also be used for evaluating their accuracy.

This procedure allows to maintain control over the errors made during model equation development.

### A Hole charge partitioning

The bias dependence of the integrand in  $\Delta Q_{ph}$  of (4.55) is difficult to describe analytically, unless one separates  $\Delta$  into its contribution to (i) charging the depletion regions and (ii) neutralizing the injected minorities. This can be done analogously to the small-signal full Regional Approach (FRA) in sec. 2.4 by defining the excess minority carrier density variation

$$\Delta m = m - m_0 = \begin{cases} \Delta p & , |\Delta p| < |\Delta n| \\ \Delta n & , |\Delta n| \leq |\Delta p| \end{cases} \quad (4.60)$$

with  $m_0$  as equilibrium density. The corresponding variation of the total weighted excess minority charge in the transistor is then given by

$$\Delta Q_{mh} = qA_E \int_0^{L_x} h \Delta m dx . \quad (4.61)$$

With the weighted excess hole charge from (4.55), the weighted depletion charge is then automatically given by

$$\Delta Q_{jh} = \Delta Q_{ph} - \Delta Q_{mh}$$

$$= qA_E \left[ \int_0^{L_x} h \Delta p dx - \left( \int_{x(|\Delta p_i| \leq |\Delta n_i|)} h \Delta p dx - \int_{x(|\Delta n_i| \leq |\Delta p_i|)} h \Delta n dx \right) \right]$$

with  $\Delta p_h$  ( $\Delta n_h$ ) defined as  $h\Delta p$  ( $h\Delta n$ ). The first and second integral cancel in regions where weighted holes are minorities, so that the weighted depletion charge is simply

$$\Delta Q_{jh} = qA_E \int_{x(|\Delta n_h| \leq |\Delta p_h|)} h(\Delta p - \Delta n) dx . \quad (4.62)$$

In contrast to the actual depletion charge, there is no exact neutrality of the weighted depletion charge in a transistor, i.e.

$$\int_0^{L_x} h(\Delta p - \Delta n) dx \neq 0 .$$

As a consequence, the weighted depletion charge can only be calculated from the excess carrier densities in the base region.

The question is whether the regions that define the actual minority carriers and their weighted counterparts are identical. This is important since according to (4.62) the calculation of the weighted depletion charge extends over a region that is defined according to (4.60) by the cross-over point  $x_{mh}$ ,

$$\{h(\Delta p - \Delta n)\} \big|_{x_{mh}} = 0.$$

A small-signal perturbation of this expression leads to

$$\{h(\delta p - \delta n)\} \big|_{x_{mh}} + \{\delta h(\Delta p - \Delta n)\} \big|_{x_{mh}} = 0.$$

Since the small-signal densities equal zero at  $x_m$  and the second term is zero,  $x_{mh} = x_m$  must hold.

The next step is to link the weighted charge components with the corresponding actual charges that can be described by measurable quantities. Of course, one can always define a weight factor  $\bar{h}$  as the ratio of a weighted charge to its actual charge counterpart. However, it may be difficult to assign a physical meaning to the weight factor this way and to derive an analytical expression.

## B Depletion charge

The weighted depletion charge can be divided into a BE and BC component:

$$\Delta Q_{jEih} = qA_E \int_{x_{me}}^{x_{Bn}} h(\Delta p - \Delta n) dx, \quad (4.63a)$$

$$\Delta Q_{jCih} = qA_E \int_{x_{Bn}}^{x_{mc}} h(\Delta p - \Delta n) dx. \quad (4.63b)$$

with  $x_{Bn}$  defined as location in the base where  $(\Delta p - \Delta n) = \min$ . Assuming that the classical theory can be applied to the large-signal space-charge density variation  $\Delta \rho = q(\Delta p - \Delta n)$ , any analytical description of the weighted charge will obviously depend on the spatial dependence of  $h$ . However, it is always possible to write

$$\Delta Q_{jEih} = \bar{h}_{BE}(V_{B'E}, V_{B'C}) q A_E \int_{x_{me}}^{x_{Bn}} (\Delta p - \Delta n) dx, \quad (4.64)$$

in which the spatially averaged weight factor  $\bar{h}_{BE}$  generally depends on bias. The remaining integral corresponds to a depletion charge

$$\Delta Q_{jEi} = q A_E \int_{x_{me}}^{x_{Bn}} (\Delta p - \Delta n) dx. \quad (4.65)$$

The question now is whether  $\Delta Q_{jEi}$  can be related to the actual depletion charge  $Q_{jEi}$  obtained from small-signal calculations or measurements. Extending the integrand of (4.65) in form of a derivative w.r.t to the BE terminal voltage and subsequent integration over the same voltage yields the equivalence

$$\Delta Q_{jEi} = q A_E \int_{x_{me}}^{x_{Bn}} \left( \int_0^{V_{B'E}} \frac{\partial(\Delta p - \Delta n)}{\partial V_{B'E}} \bigg|_{V_{B'C}} dV \right) dx. \quad (4.66)$$

The integration limit  $x_{Bn}$  does not depend on bias, while the boundary  $x_{me}$  may have very small bias dependence. Neglecting the latter makes the spatial and bias integration limits independent of each other, thus allowing to switch the integration variables and associated sequence:

$$\Delta Q_{jEi} = \int_0^{V_{B'E}} \left( q A_E \int_{x_{me}}^{x_{Bn}} \frac{\partial(\Delta p - \Delta n)}{\partial V_{B'E}} \bigg|_{V_{B'C}} dx \right) dV. \quad (4.67)$$

The relation between the excess and the small-signal carrier densities in the BE SCR is visualized in Fig. 4.23. Noticing that the equilibrium densities  $p_0$  and  $n_0$  in  $\Delta p$  and  $\Delta n$  do not depend on bias, i.e.  $\partial \Delta p = \partial p$  and  $\partial \Delta n = \partial n$ , and due to charge neutrality used in writing (2.22) gives for the expression in the parenthesis the same as employed in (2.43) for the BE depletion capacitance from the FRA. As a consequence,

$$\Delta Q_{jEi} = \int_0^{V_{B'E}} C_{jEi} dV = Q_{jEi}. \quad (4.68)$$

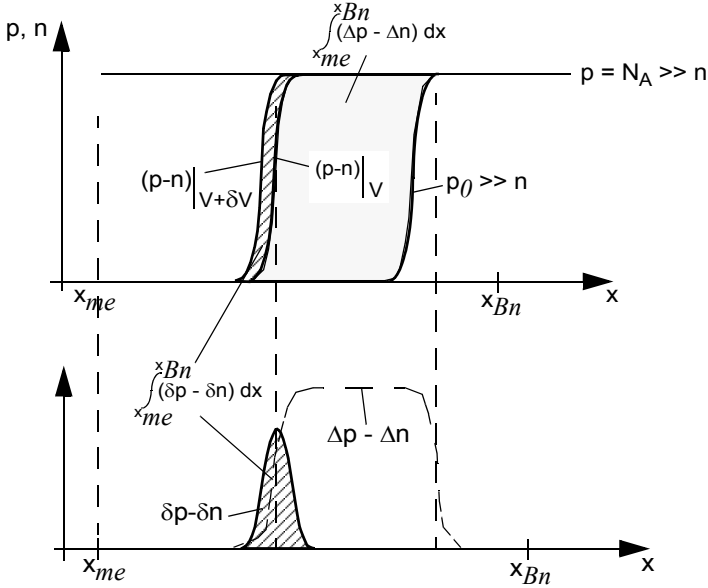


Fig. 4.23: Illustration of the equivalence of the carrier and associated charge variation between the large-signal case (upper figure) and small-signal case (lower figure). A spatially constant profile is assumed.

Even if the crossover point  $x_{me}$  changes somewhat with bias the error in the integration remains negligible, since the space charge variation around this point is very small compared to the variation further away from the crossover.

The equivalence between (4.65) and (4.68) has also been confirmed by device simulation. The relation between the weighted and actual BE depletion charge then reads

$$\Delta Q_{jEih} = \bar{h}_{BE} Q_{jEi}. \quad (4.69)$$

In other words, the bias dependent weighted GICCR depletion charge can be expressed by the actual depletion charge that is measurable via the transistor terminals. Hence, (4.69) can be used as a definition of  $\bar{h}_{BE}$ .

Figure 4.24 shows the bias dependence of the weight factor. For BJTs  $\bar{h}_{BE}$  is smaller than 1 in accordance with Fig. 4.21a, and the bias dependence is fairly small due to the weak spatial dependence of  $h(x)$  in the BE SCR (cf. Fig. 4.17). In contrast, for the HBT  $\bar{h}_{BE}$  is significantly larger



than 1 and strongly bias dependent since the Ge profile shape causes the bandgap in the BE SCR not only to be larger than in the neutral base region but also to be strongly spatially dependent.

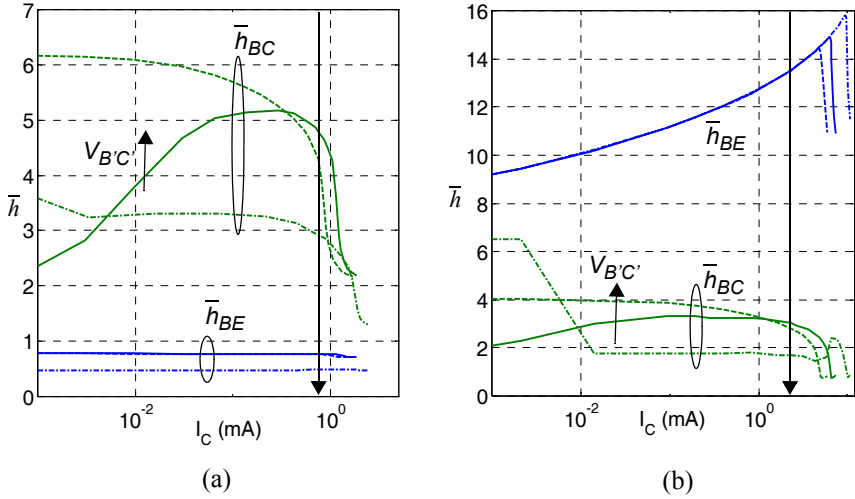


Fig. 4.24: Bias dependence of the average weight factors  $\bar{h}_{BE}$  and  $\bar{h}_{BC}$  for the depletion charges obtained from device simulation: (a) BJT, (b) CED HBT.  $V_{B'C'}/V = 0.4$ , 0, -2.3. The arrow indicates peak  $f_T$ .

Applying the same approach to the BC depletion charge leads to the equivalence

$$\Delta Q_{BCi} = qA_E \int_{x_{mc, \Delta}}^{L_x} (\Delta p - \Delta n) dx = Q_{BCi}. \quad (4.70)$$

with

$$Q_{BCi} = \int_0^{V_{B'C}} C_{jCi} dV + \int_0^{I_T} \tau_{BC} dI,$$

which includes the modulation of the BC space charge with a  $V_{B'E'}$  change. Hence, at a given bias point one can write

$$\Delta Q_{BCih} = \bar{h}_{BC}(V_{B'E'}, V_{B'C}) Q_{BCi}, \quad (4.71)$$

which links the weighted depletion charge with its actual counterpart through a spatially averaged weight factor  $\bar{h}_{BC}$ . As exhibited in Fig. 4.24 the bias dependence of  $\bar{h}_{BC}$  is similar for BJT and HBT. The variation with

$V_{B'C'}$  results from the field dependence of the mobility and its impact on  $h(x)$  in the BC SCR.

### C Minority charge

Again, the goal is to relate the weighted minority charge variation (4.61) in a physics-based way to the actual minority charge or its components, for which compact analytical approximations already exist. For the time being, only the neutral base region shall be considered, the weighted charge of which reads

$$\Delta Q_{mBh} = qA_E \int_{x_e}^{x_c} h \Delta n dx. \quad (4.72)$$

The question is whether the excess minority charge

$$\Delta Q_{mB} = qA_E \int_{x_e}^{x_c} \Delta n dx \quad (4.73)$$

can be linked to the corresponding expression that results from integrating a storage time, which in this case is

$$Q_{nB} = \int_0^{I_T} \tau_{mB} dI. \quad (4.74)$$

For the simple case of a diffusion transistor, the development of  $\Delta n(x)$  with bias and its relation to the small-signal change is illustrated in Fig. 4.25.

Replacing in  $\Delta Q_{mB}$  the carrier density at each location  $x'$  (cf. Fig. 4.25) by the summation of its incremental changes with bias from equilibrium to  $I_T$  gives

$$\Delta Q_{mB} = qA_E \int_{x_e}^{x_c} \left( \int_0^{I_T} \frac{\partial \Delta n}{\partial I_T} \bigg|_{V_{CE}} dI \right) dx. \quad (4.75)$$

This is equivalent to integrating the small signal charges (hatched area in Fig. 4.25) at each bias point from equilibrium to  $I_T$ , i.e.

$$\Delta Q_{mB} = \int_0^{I_T} \left( qA_E \int_{x_e}^{x_c} \frac{\partial \Delta n}{\partial I_T} \bigg|_{V_{CE}} dx \right) dI. \quad (4.76)$$

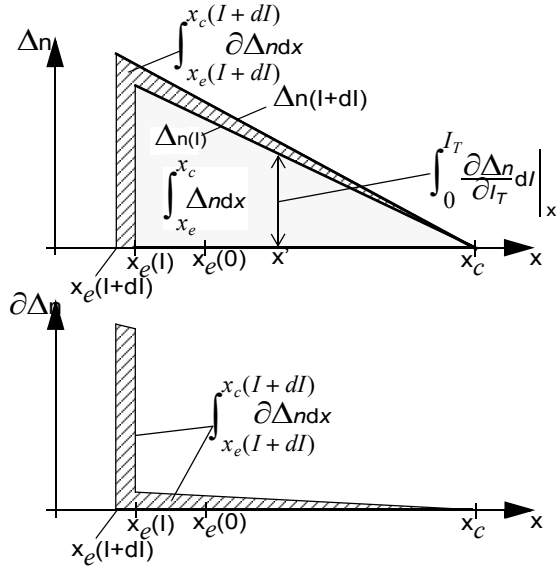


Fig. 4.25: (a) Sketch of excess electron density and the corresponding minority charge in the neutral base at a given bias point ( $I = I_T$ ) and for a change from  $I_T$  to  $I_T + dI_T$ . (b) Sketch of corresponding small-signal electron density.

According to (2.38) and with  $\partial \Delta n = \partial m$  in the neutral base, the inner integral corresponds to the base storage time at a given bias point. Hence,

$$\Delta Q_{mB} = \int_0^{I_T} \tau_{mB} dI = Q_{nB}, \quad (4.77)$$

and  $\Delta Q_{mBh}$  in (4.72) can be related to the actual base minority charge,

$$\Delta Q_{mBh} = \bar{h}_{mB} Q_{nB}, \quad (4.78)$$

in which the corresponding weight factor  $\bar{h}_{mB}$  has a physical meaning and may be bias dependent. The equivalence between (4.77) and (4.73) has been verified by device simulation.

In a similar way, the other components of the weighted minority charge can be linked to their corresponding actual minority charges:

$$\Delta Q_{mEh} = \bar{h}_{mE} Q_{mE}, \quad (4.79)$$

$$\Delta Q_{mBEh} = \bar{h}_{mBE} Q_{mBE}, \quad (4.80)$$

$$\Delta Q_{mBCh} = \bar{h}_{mBC} Q_{mBC}, \quad (4.81)$$

$$\Delta Q_{mCh} = \bar{h}_{mC} Q_{mC}. \quad (4.82)$$

The actual charges on the r.h.s. are given by integrating the regional storage times over the transfer current. Note, that (4.81) only holds if minority carriers are inserted; i.e the electrons in the BC SCR must not be included. The relations above define the weight factors. Since the latter are spatial averages they may be bias dependent. A first-order approximation is to assume bias independent values, e.g. averaged over the relevant bias range. This has worked fairly well in existing compact models but may not yield sufficiently accurate results any more for future technologies.

In practice, not all of the minority charge components (4.78)-(4.82) are explicitly taken into account in a compact model. For instance, the minority charge  $Q_{mBE}$  is usually neglected. Also, the minority (hole) charge on the collector side of the BC SCR is included in  $Q_{pC}$  while its counterpart (electrons) on the base side is neglected, so that  $Q_{mBC}$  is dropped. The errors made by these simplifications are usually compensated for by the adjustment of model parameters and weight factors of other components.

#### D Compact formulation

Assembling the various charge components including the zero-bias charge from (4.57) gives with (4.56)

$$Q_{ph} = \bar{h}_0 Q_{p0} + \bar{h}_{BE} Q_{jEi} + \bar{h}_{BC} Q_{jCi} + \sum_k \bar{h}_{mk} \Delta Q_{mk} \quad (4.83)$$

with  $k = \{E, BE, B, BC, C\}$  indicating the regional minority charge components. Division by  $\bar{h}_0$  allows to eliminate one factor (and potential model parameter), leading to the transfer current related hole charge

$$Q_{pT} = \frac{Q_{ph}}{\bar{h}_0} = Q_{p0} + h_{jEi} Q_{jEi} + h_{jCi} Q_{jCi} + \sum_k h_{fk} \Delta Q_{mk} \quad (4.84)$$

with the normalized weight factors

$$h_{jEi} = \frac{\bar{h}_{BE}}{\bar{h}_0}, \quad h_{jCi} = \frac{\bar{h}_{BC}}{\bar{h}_0}, \quad h_{fk} = \frac{\bar{h}_{mk}}{\bar{h}_0}. \quad (4.85)$$

As a consequence of the normalization, the prefactor in (4.52) changes to

$$c_1 = A_E q \frac{c_0}{\bar{h}_0} = (A_E q)^2 V_T \overline{\mu_{nB0} n_{iB}^2}. \quad (4.86)$$

with the last term taken at equilibrium. Inserting the above variables into (4.52) yields

$$I_T = c_1 \frac{\exp\left(\frac{V_{B'E'}}{V_T}\right) - \exp\left(\frac{V_{B'C}}{V_T}\right)}{Q_{pT}}, \quad (4.87)$$

which is a more suitable basis for HBT compact modeling and is referred to as generalized integral charge-control relation (GICCR), which contains the ICCR as a subset [28,17,18].

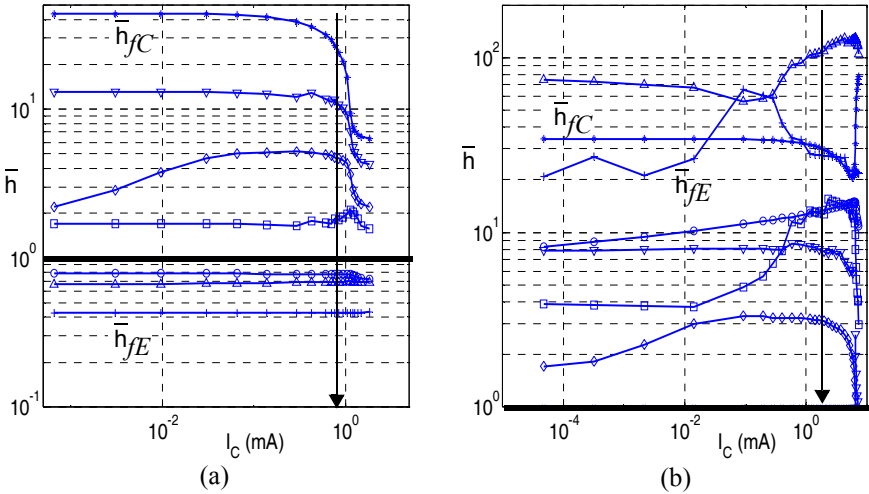


Fig. 4.26: Current dependence of the normalized weight factors for the various charge components, obtained from device simulation for  $V_{B'C} = 0V$ : (a) BJT and (b) CED HBT. Legend:  $\bar{h}_{fEi}(o)$ ,  $\bar{h}_{fCi}(\square)$ ,  $\bar{h}_{fE}(+)$ ,  $\bar{h}_{fBE}(\nabla)$ ,  $\bar{h}_{fB}(\diamond)$ ,  $\bar{h}_{fBC}(\Delta)$ ,  $\bar{h}_{fC}(*).$  The arrow indicates peak  $f_T$ .

Depending on the doping profile, the normalized weight factors may be a function of bias as shown in Fig. 4.26. For the BJT, most factors are fairly constant up to peak  $f_T$ , except  $\bar{h}_{fB}$ . However, the variation of the latter at low current densities has little impact, since the minority charge is very

small there. Although for the HBT more factors are varying at lower current densities, only the bias dependence of  $\bar{h}_{jEi}$  is relevant since  $Q_{jEi}$  is the largest contribution in this bias range. At high current densities several factors vary quite strongly but the relevance to compact modeling is limited. The wiggles in some of the curves result from a more difficult regional evaluation due to the quite rapid Ge profile variation at the junctions.

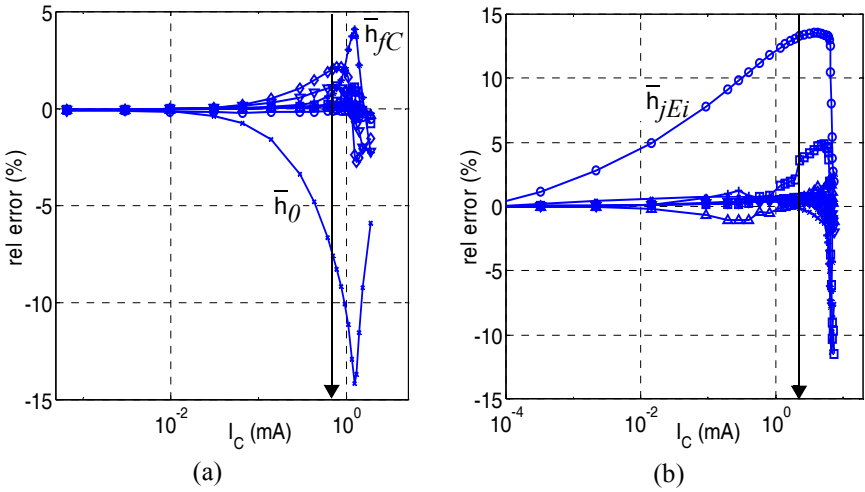


Fig. 4.27: Relative error of transfer current from the GICCR with the weight factors and charge components taken directly from device simulation using the regional approach: (a) BJT, (b) CED HBT.  $V_{B^+C^+} = 0V$ . In each curve just *one* weight factor has been set constant:  $\bar{h}_{jEi}(\circ)$ ,  $\bar{h}_{jCi}(\square)$ ,  $\bar{h}_{jE}(+)$ ,  $\bar{h}_{jBE}(\nabla)$ ,  $\bar{h}_{jFB}(\diamond)$ ,  $\bar{h}_{jBC}(\triangle)$ ,  $\bar{h}_{jC}(\ast)$ . The arrow indicates peak  $f_T$ . For this evaluation, the non-normalized form of the GICCR with  $c_0$  has been used.

Note, that (4.87) is still exact as long as the exact charges and weight factors are inserted (e.g. from device simulation). The question is how large an error is introduced if the exact charge is multiplied with just a constant weight factor. Figure 4.27 contains an overview on the relative error in calculating  $I_T$  from (4.87) with one charge component being replaced by the product of the actual charge with constant weight factor, while all others remain exact. In most cases the error is  $\pm 5\%$  or even well below. Only the impact of  $\bar{h}_0$  for the BJT and  $\bar{h}_{jEi}$  for the HBT is significantly larger, reaching about 15% at high current densities. The variation of  $\bar{h}_0$  and, thus,  $Q_{p0h}$  with bias is less visible in  $I_C(V_{BE})$  characteristics, but influences mostly

the transconductance of the BJT. The effect is similar for the HBT due to the variation of  $\bar{h}_{jEi}$  at constant  $Q_{p0h}$ . Overall, using constant factors for all components will lead even to some error compensation. Notice, that normalizing to a bias dependent  $\bar{h}_0$  will cause factors of the other components and also  $c_I$  to become bias dependent. Hence, for the BJT the form (4.87) is less advantageous than for the HBT. The considerations so far have been restricted to the 1D case in order to demonstrate the concept, but the GIC-CR can be extended to the 3D case (cf. [18] and sec. 5.3.2).

According to section 4.1.3 the graded Ge contents causes holes to shift from the emitter end of the base region to the collector end of the base, which is the reason for the aiding drift field. The corresponding increase of hole density at the collector leads to some compensation of ionized acceptors and, hence, a reduction of the BC SCR width in the base compared to the case with spatially constant Ge contents. As a consequence, the Early effect in transistors with graded Ge is reduced. On the other hand, the missing holes at the emitter end lead to a significantly larger “reverse” Early-effect for a graded Ge profile. This can also be shown analytically by evaluating the GICCR with sufficiently simple doping and Ge profiles.

### 4.3.2 Base current components related to recombination

At medium injection levels under forward operation the main component of the base current is still given by emitter backinjection. In advanced transistors recombination in the poly-silicon emitter dominates even more. This is due to (i) the extremely reduced width of the mono-silicon emitter region (on the order of 10nm or less) and (ii) processing techniques that eliminate the thin oxide layer at the poly-mono interface and related recombination there. The elimination of the oxide layer is mandatory also for reducing the emitter resistance in order to achieve high operating speed.

Fundamentally the theory discussed in sec. 3.3.2 applies. The question though is whether a possibly existing bandgap difference due to a variation in material composition and its possible shift into the neutral emitter has an influence on the base current  $I_{jBEi}$ . Such a shift can occur for fabrication reasons in order to safely avoid any barrier for electrons in the (neutral) base. Figure 4.28 illustrates the situation, in which the neutral emitter por-





Replacing  $w_E$  by  $x_{he}$  in the mono-Si expressions of sec. 3.3.2 and inserting the resulting equations together with the boundary and interface conditions yields for the back-injection current

$$I_{jBEi} = \frac{qA_E \frac{\bar{\mu}_{pEh} V_T}{w_E} \frac{n_{iE, SiGe}^2(w_E)}{N_E(w_E)} \left[ \exp\left(\frac{V_{B'E'}}{V_T}\right) - 1 \right]}{\frac{w_E - x_{he}}{w_E} + \left( \frac{\bar{\mu}_{pEh} x_{he}}{\bar{\mu}_{pE} w_E} + \frac{\bar{\mu}_{pEh} (V_T/w_E)}{v_{Epm}} \right) \exp\left(\frac{\Delta V_V}{V_T}\right)}. \quad (4.91)$$

For  $x_{he} = w_E$  the mono-Si solution (3.99) is recovered. A more detailed analysis for the case  $x_{he} < w_E$  reveals that the difference in current between (4.91) and (3.99) is very small if one assumes that the bandgap difference occurs completely in the valence band since then the enhancement resulting from  $n_{iE, SiGe}^2$  is mostly compensated by the  $\exp(\Delta V_V/V_T)$  term in the denominator. This difference is even smaller for often used triangular Ge profiles (cf. curve 1 in Fig. 4.6). As a consequence, (3.101) can still be used even if during forward operation of the BE junction the heterojunction becomes exposed in the neutral emitter.

Fundamentally, recombination in the neutral base volume exists in both BJTs and HBTs. While neutral base recombination (NBR) has turned out to be negligible in modern BJTs, the lower bandgap in the base of SiGe HBTs may lead to an increased level of NBR. In general, the effect can be captured analytically by solving the diffusion equation for minorities in the neutral base, which can be found in standard text books. Subtracting then the current reaching the collector side (given by the base transport factor) from the current injected at the emitter side gives for npn transistors

$$I_{NBR} = I_{NBRs} \left[ \exp\left(\frac{V_{B'E'}}{V_T}\right) - 1 \right] \quad (4.92a)$$

with the saturation current

$$I_{NBRs} \approx qA_E \frac{\bar{\mu}_{nB} V_T}{\bar{L}_{nB}} \frac{\bar{n}_{iB}^2}{\bar{N}_B} \frac{\cosh(w_B/\bar{L}_{nB}) - 1}{\sinh(w_B/\bar{L}_{nB})}. \quad (4.92b)$$

The “≈” sign results from taking average values for the base doping and intrinsic carrier density, which are different from those for each of the above

mentioned current components in case of a spatially varying doping or Ge profile.

For negligible avalanche current, it is evident from (4.92b) that the impact of NBR may be observed through a variation of the base width  $w_B$  with  $V_{B'C'}$ , since the emitter backinjection component does not depend on  $V_{B'C'}$ . The magnitude of the variation depends strongly on the Ge profile and processing conditions and has been debated in literature (e.g. [29, 30]). Existing data indicates though that the effect is very small at room temperature and above, and only seems to be observable at temperatures well below those specified for typical circuit applications. The effect can be taken into account in a compact model using a suitable description for the base width variation with  $V_{B'C'}$ .

In SiGe HBTs significant additional recombination can occur at high current densities. In contrast to BJTs, the valence band barrier at the BC junction prevents holes from being injected into the collector and instead causes them pile up at the barrier. The resulting dipole effect then starts to also create a barrier for electrons. The combined pile-up of both carrier types in front of the barrier then causes excess recombination, which becomes visible in the terminal base current as additional increase with  $V_{B'E'}$ . The recombination must be proportional to the excess base minority charge  $\Delta Q_{Bn}$ . Denoting  $\tau_{Bhrec}$  as the corresponding (average) lifetime, the effect can be described by the simple relation [31, 32]

$$I_{Bhrec} = \frac{\Delta Q_{Bn}}{\tau_{Bhrec}}. \quad (4.93)$$

Note that this additional current component generally is observed only at *very high* collector current densities, at which the transistor operating speed has already significantly decreased (e.g. [32, 33]) and severe self-heating occurs. Hence, the relevance of  $I_{Bhrec}$  for circuit design and compact modeling is quite limited.

#### 4.3.3 Non-local base-collector avalanche current

With shrinking vertical device dimensions and with - at the same time - increased doping concentrations, breakdown phenomena become an increasingly important issue for integrated circuit design and device reliabil-

ity. At the same applied voltage, higher doping concentrations lead to a decrease of the SCR width while the electric field increases. In other words, the width of the high-field region that is most relevant for the ionization integral (3.123) decreases with each new technology generation. In today's SiGe HBTs, the SCR width already corresponds to just a few times the energy relaxation (dissipation) length. The latter determines the distance over which carriers gain kinetic energy before they are able to cause impact ionization. Therefore, using the local electric field, as in sec. 3.3.4, yields a too large ionization integral and avalanche current. This was first observed in [34] and is shown in Fig. 4.29. From low to high voltages the multiplication factor calculated from the local field varies over a little more than a decade due to the weak (square root) dependence of the maximum field on voltage and due to the fact that only a small region around that peak contributes to the ionization integral. In contrast, the non-local multiplication factor varies over several decades and is much lower even at high voltages, i.e. high fields. This is due to the finite distance required for accumulating sufficient kinetic energy, which corresponds to a lower effective field with a wider peak region.

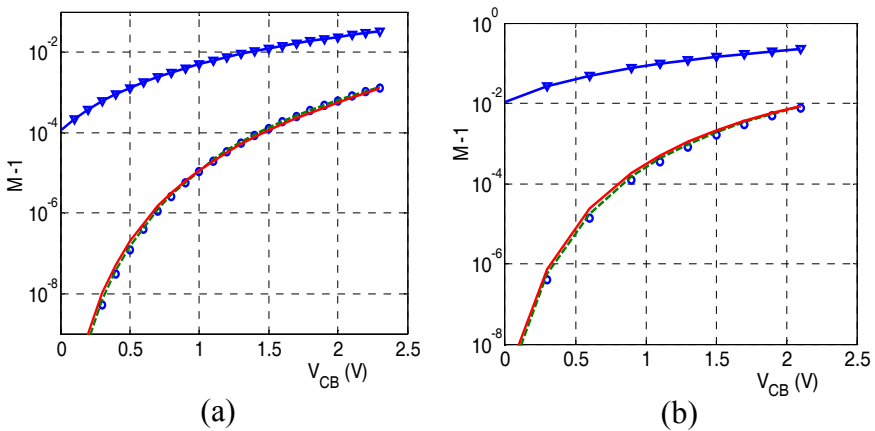


Fig. 4.29: Avalanche multiplication factor vs. reverse BC voltage for the local field model (line with triangle), the non-local field model from device simulation (o), from (4.100) (dashed line), and from (4.101) (solid line): (a) BJT, (b) CED HBT.

The “delay” in accumulating energy when being accelerated by an electric field can be described to first order by an energy balance (EB) equa-

tion, which results from taking the second moment of the Boltzmann transport equation. As shown in [34, 35] a quite accurate description of the multiplication factor can be obtained from the simplified form (2.13) of the EB equation if the local electric field in the ionization integral (3.123) is replaced by an effective field that can be defined as

$$E_{n,av}(x) = \frac{5k_B T_n(x) - T_L}{2q} \frac{1}{\lambda_n} = \frac{1}{\lambda_n} \int_0^x E_n(\xi) \exp\left(\frac{\xi - x}{\lambda_n}\right) d\xi. \quad (4.94)$$

This relation follows from the steady-state ( $dT/dx = 0$ ) solution of (2.13) and is calculated from the solution (2.14) of the EB equation.

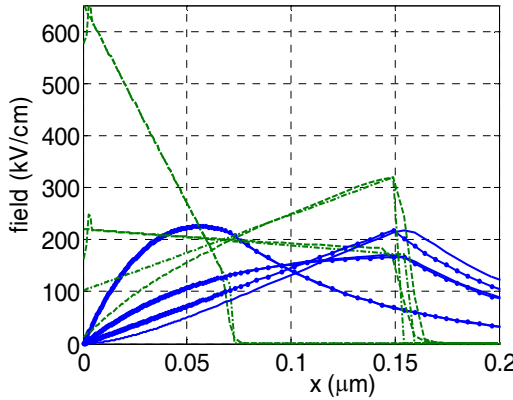


Fig. 4.30: Local field calculated from the gradient of the electron QFP (dashed lines) and resulting non-local effective carrier field (solid lines, using  $\lambda_n = 70\text{nm}$ ) in the collector region for the CED HBT at the bias points  $(V_{B'E}, I_C/\text{mA}) = (0.70, 0.0022), (0.93, 7.5), (0.94, 9.7)$  and  $V_{B'C} = -2.3\text{V}$ . For comparison, also the analytical local field model (4.95) (dash-dotted lines) and the resulting effective field (4.97) (lines with dots) are inserted.

Figure 4.30 contains a comparison between the magnitude of the actual and the effective field in the collector region (starting with  $x = 0$  at the BC junction). Note that the field  $E_n$  relevant for the avalanche effect is given by the gradient of the respective carrier quasi-Fermi potential and not the electrostatic potential. At low current densities, the local field assumes the expected triangular shape until the end of the SCR. However, in contrast to the *electrostatic* field,  $E_n$  drops sharply to a low value at the SCR boundary. This causes the effective field to be quite non-symmetric around its

peak, which is close to the SCR boundary. The convolution of  $E_n$  with the exponential weight function leads to a significant reduction of the effective field, which is directly proportional to the carrier energy and temperature; in other words, the carrier temperature and, thus, the velocity not only are much lower compared to the local case but also can peak at a location far within the SCR. Towards high current densities, the electrostatic field starts to change its slope, and the SCR extends to the buried layer. This causes the peak of the effective field to shift to the end of the collector. Also its shape changes to a more symmetrical form. Note that in the non-local case the carriers enter the buried layer with still significant energy.

A compact model requires an analytical expression for the ionization integral. For this, the extension of the BC SCR into the base is neglected and, according to Fig. 4.30, a triangular shape for the *magnitude* of the local field in the BC portion of the SCR is assumed,

$$E_n(x) = \begin{cases} E_{jc} - s_E \cdot x & , 0 \leq x < w_{BC} \\ 0 & , w_{BC} \leq x \leq w_{Ci} \end{cases} , \quad (4.95)$$

with  $E_{jc}(V_{B'C'}, I_T)$  as bias dependent (maximum) field at the junction. This approach was also pursued in [36], where the low-injection case was assumed so that  $s_E = E_{jc}(V_{B'C'}, 0)/w_{BC}$ . In the general case, the slope is also determined by the current and reads

$$s_E = \frac{qN_{Ci}}{\varepsilon} \left( 1 - \frac{I_T}{I_{lim}} \right). \quad (4.96)$$

As Fig. 4.30 shows, the above approximation describes the results of device simulation quite well for all current densities. In a compact model,  $w_{BC}$  may be calculated from the BC depletion capacitance according to (3.132).

Evaluating the integral in (4.94) yields for the effective field

$$E_{n, av}(x) = \begin{cases} (E_{jc} + \lambda_c s_E) \left[ 1 - \exp\left(\frac{-x}{\lambda_c}\right) \right] - s_E x & , 0 \leq x \leq w_{BC} \\ E_{n, av}(w_{BC}) \exp\left(\frac{w_{BC} - x}{\lambda_c}\right) & , w_{BC} < x \leq w_{Ci} \end{cases} \quad (4.97)$$

where  $E_{n,av}(w_{BC})$  is calculated from the upper equation. According to Fig. 4.30, the above expression is an excellent compact approximation of the effective field. In fact, the convolution significantly reduces the deviations of the local field model (4.95) and, thus, is advantageous for compact modeling.

Expression (4.97) does not allow to evaluate the ionization integral (3.123) analytically. Thus, the argument of the exponent in (3.123),

$$b_{av}(x) = \frac{b_n}{E_{n,av}(x)}, \quad (4.98)$$

needs to be approximated by a suitable function. As Fig. 4.31 shows,  $b_{av}$  is fairly symmetric at low current densities and for a partially depleted collector, but becomes quite non-symmetric for a fully depleted collector, which occurs at either high voltages or high current densities. As can also be observed, the model (4.98) with (4.97) is quite a good approximation.

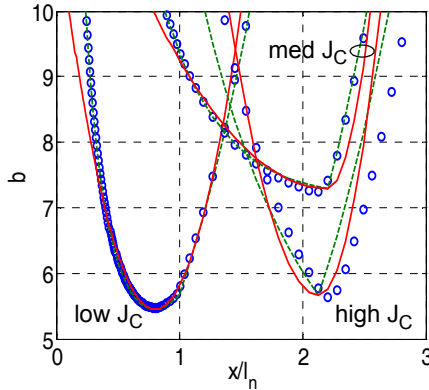


Fig. 4.31: Ionization function  $b_{av}$ , defined by (4.98), from device simulation (symbols), analytical effective field model (dashed lines), and analytical approximation (4.99) (solid lines) for the CED HBT at bias points  $(V_{BE}/V, I_C/\text{mA}) = (0.70, 0.0022)$ ,  $(0.93, 7.5)$ ,  $(0.94, 9.7)$ .  $V_{BC} = -2.3V$ .

The ionization integral value is determined by the region around the minimum  $b_{avm}$  up to roughly  $b_{avm} + 2$ , at which the integrand values have already decayed by the factor  $e^2$ . Therefore, as approximation of  $b_{av}$ , a simple quadratic function is used,

$$b(x') = b_{avm} + \begin{cases} \left( \frac{x' - x_{avm}}{l_{avl}} \right)^2 & , x' \leq x_{avm} \\ \left( \frac{x' - x_{avm}}{l_{avr}} \right)^2 & , x' > x_{avm} \end{cases} , \quad (4.99)$$

with  $x' = x/\lambda_n$  as normalized spatial variable. The function is continuous and assumes its minimum  $b_{avm}$  at the point  $x_{avm}$ , at which both functions are linked together. The two functions only differ in shape through the different parameter  $l_{av(l,r)}$  for the left- and right-hand-side. The parameters can be determined from evaluating (4.98) with (4.97) at four points  $x'$ . Note that this function does not necessarily have its minimum at the same location where a maximum of the *upper* part of  $E_{n,av}$  in (4.97) occurs; in fact the latter does not have a maximum at all at sufficiently high current densities. As it turned out, the best method for calculating the parameters is to select three points at  $x'_{(1,2,3)} = x_{avm} - (0, \Delta_l/2, \Delta_l)$  with  $\Delta_l$  as the only adjustable value (typical range  $\Delta_l = 0.3 \dots 1$ ) and one point at  $x'_4 = x_{avm} + \Delta_l/2$ . The resulting approximation is also inserted in Fig. 4.31 for selected current densities and exhibits a reasonably good agreement especially at low current densities.

With the analytical approximation, the multiplication factor (3.123) now reads

$$M_n \cong 1 + a_n \left[ \int_0^{x_{avm}} \exp \left( - \left[ b_{avm} + \left( \frac{\xi - x_{avm}}{l_{avl}} \right)^2 \right] \right) d\xi \right. \\ \left. + \int_{x_{avm}}^{x_C} \exp \left( - \left[ b_{avm} + \left( \frac{\xi - x_{avm}}{l_{avr}} \right)^2 \right] \right) d\xi \right] \quad (4.100)$$

with  $x_C$  as the 1D collector contact point in the buried layer. Contributions to the integral only come from a region around the peak of the effective field, i.e. the minimum of  $b(x')$ . Once  $b(x')$  has increased from its minimum by about 2 to 3 the contributions very quickly fade. Therefore, the lower and upper integration limits 0 and  $x_C$ , respectively, can be expanded to infinity without changing the value of the integral. This leads to the compact expression:

$$M_n \cong 1 + a_n \sqrt{\pi} \frac{(l_{avl} + l_{avr}) \lambda_n}{2} \exp(-b_{avm}). \quad (4.101)$$

Figure 4.32 shows the multiplication factor from device simulation using the non-local avalanche model. The factor first increases slightly with current due to the increasing SCR width and the widening of the effective field towards the buried layer. In contrast, with a local field model the decrease of  $E_{jc}$  with current would lead to a decrease of  $M$  due the direct dependence on  $E_{jc}$ . Eventually, this decrease also happens for the non-local model once the peak of  $E_{n,av}$  starts to drop. This trend continues until a horizontal local field is reached leading to a significant difference between low- and high-current avalanche factor. Further increase of the current leads to the formation of a local-field peak at the end of the collector and also to a reversal of the drop in effective field and  $M$ . The overall behavior is represented very accurately by the compact formulation (4.101) using the analytical field description (4.97) and the approximation (4.99). For comparison, also the results of the numerical integration of (4.98) are included.

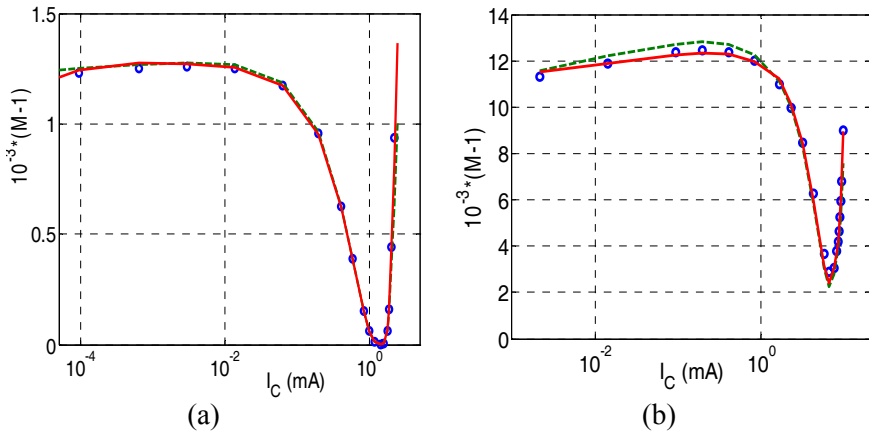


Fig. 4.32: Non-local multiplication factor from device simulation (symbols), numerical integration of ionization integral with the analytical effective field from (4.97) (dashed line), and fully analytical approximation (4.101) (solid line): (a) BJT with  $\Delta\lambda = 1.2$ ; (b) CED HBT with  $\Delta\lambda = 0.8$ .  $V_{B'C'} = -2.3\text{V}$ .

Finally, inserting  $M_n$  from (4.101) into (3.122) gives the avalanche current.



#### 4.3.4 Tunnelling

Tunnelling is a quantum mechanical effect that under certain conditions allows carriers to pass through an energy barrier although their energy is lower than the barrier height. The analysis is based on solving the Schrödinger equation which describes carrier motion in terms of waves and carrier location by the squared magnitude of the corresponding wave functions. For a detailed treatment of tunnelling the reader is referred to the literature [40, 41].

So far, tunnelling has not been an important mechanism in Si-based bipolar transistors, at least under forward operation mode. Since tunnelling at forward bias corresponds to increased leakage and reduced current gain, it is undesired and attempted to be avoided by proper device design. This is one reason (besides the BE depletion capacitance reduction) for spacing the base and emitter profile apart or reducing their slope at the junction - a measure that can be observed in all advanced process technologies. The ultimate profile leads to the LEC type SiGe HBT (e.g. [43]). Thus, mechanisms such as intra-band and trap-assisted tunnelling that occur in forward biased junctions are neglected here. In future extremely scaled HBTs though the impact of tunnelling at forward bias may be unavoidable.

The consequence of scaling the vertical dimensions during the evolution of Si-based bipolar transistors has been a constant reduction of the BE “breakdown” voltage to values below 1V in modern transistors. The corresponding increase in reverse bias leakage current needs to be taken into account in certain circuit applications. For instance, in some processes the BE junction is used as varactor diode, the quality factor of which depends on the leakage current and its associated conductance. Also, currents caused by negative BE voltage swings sometimes need to be properly accounted for in simulation. The leakage mechanism at such low reverse voltages has been originally known as *Zener* breakdown. Its physical origin is band-to-band (BTB) tunnelling. Note that - in contrast to avalanche breakdown - tunnelling is not a carrier generating process with a possible positive feedback mechanism (depending on the biasing scheme).

Assuming a sufficiently slowly varying potential with distance in the reverse biased BE junction, the BTB tunnelling current can be modeled based on the Wenzel-Kramer-Brillouin (WKB) approximation [37,38] by

$$I_{btb} = A_E \frac{\sqrt{\frac{2qm^*}{V_g}} q^2 (-V_{B'E'})}{h^2} E_j \exp\left(-\frac{8\pi\sqrt{2m^*}qV_g^{3/2}}{3hE_j}\right) \quad (4.102)$$

for  $V_{E'B'} > 0$ . Here,  $V_g (=W_g/q)$  is the (average) bandgap voltage and  $E_j$  the (average) electric field within the SCR,  $m^*$  is the effective mass, and  $h$  is Planck's constant.

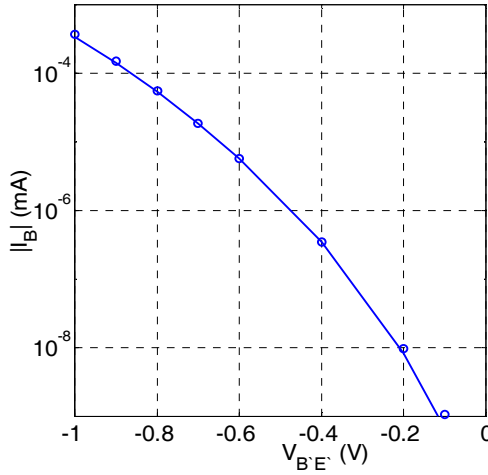


Fig. 4.33: Reverse base current from device simulation of a CED HBT with a BTB tunneling model (symbols) and analytical approximation (4.102) with (4.103) (solid lines).

$E_j$  is proportional to the maximum field, which in turn can be approximated by the theory of abrupt junctions as

$$E_j = 2 \frac{V_{DEi} - V_{B'E'}}{w_{BE}(V_{B'E'})}, \quad (4.103)$$

with  $V_{DEi}$  as built-in voltage and  $w_{BE}$  as BE SCR width. The latter can be expressed by the bias dependent BE depletion capacitance,  $w_{BE} = \epsilon / C_{jEi}$ . Inserting  $E_j$  back into (4.102) yields a compact equation for the tunnelling current that was shown in [44] to work very well for a reverse biased junction. The details of the formulation, e.g. avoiding the numerical overflow at  $V_{B'E'} = V_{DEi}$  depend on the particular compact model.

Figure 4.33 shows the reverse base current as a function of  $V_B'E'$ , calculated from device simulation with a physical BTB model. Classical theory predicts a constant saturation current at around  $10^{-15}$  mA. The observed significant increase of  $I_B$  is due to tunnelling. The compact approximation (4.102) has also been inserted in Fig. 4.33 for comparison. Using for the average field  $E_j$  simply (4.103) with  $w_{BE} = \varepsilon/C_{jEi}$  yields excellent agreement.

## 4.4 Charge storage

The impact of material composition and high-current effects on charge storage will be discussed along with the associated extensions of classical theory. Charge storage determines the dynamic transistor behavior. Its modeling is important for obtaining the capacitances and storage times. Only the latter can be directly measured via the terminals.

### 4.4.1 Depletion charges

Classical theory with the extensions presented in sec. 3.4 has worked quite well in standard compact models, even for HBTs. This aspect should be discussed though in more detail to obtain a better feeling for possible validity limits and the relation to physical and technology data. The corresponding insight will at least be helpful for statistical modeling. Furthermore, the current dependence of especially the BC depletion capacitance has been completely ignored so far. However, the associated nonlinearity with bias can be important in applications that are sensitive to distortion.

#### 4.4.1.1 Influence of the material composition

The heterojunction and its location have certain consequences for the depletion charge and capacitance. Figure 4.34 visualizes the impact a heterojunction and its location have on the BC depletion capacitance for the example of the CED HBT. Curve 1 corresponds to the Ge profile used so far. There is no fundamental difference in shape to the  $C(V)$  relation (3.139) observed for the corresponding Si homojunction (curve 3), except

for a small offset in absolute values. However, if the heterojunction is moved away from the pn-junction significantly into the collector (curve 2), a kink can be observed at forward bias. These cases will be considered below in more detail.

The theoretical consideration of a pn-junction with varying material composition is based on Poisson's equation. Assuming a spatially constant doping profile with the pn-junction located at  $x = 0$  and neglecting any mobile carriers in the SCR with width  $w$ , Poisson's equation (2.2) reads for a 1D structure

$$\frac{d[\varepsilon(x)E(x)]}{dx} = \rho, \quad (4.104)$$

in which the permittivity  $\varepsilon(x)$  is a function of the material composition.

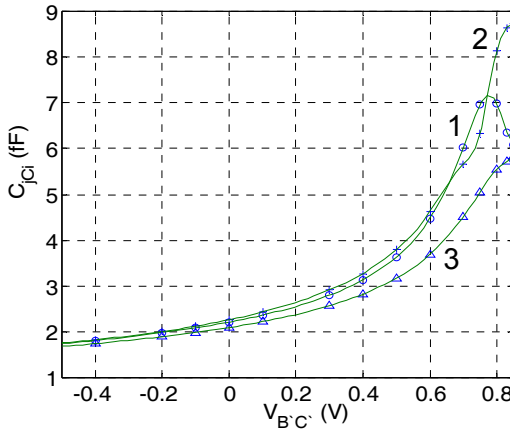


Fig. 4.34: Voltage dependence of the BC depletion capacitance for the CED HBT (curve 1), the same structure with the heterojunction moved by 30nm in the collector (curve 2), the homojunction version of the CED HBT (i.e. no Ge, curve 3).

To start with, also an abrupt change in material composition is assumed that is exactly aligned with the pn-junction as sketched in Fig. 4.35a. Solving Poisson's equation basically the same way as for a homojunction yields the same relation (3.139) for the voltage dependence as classical theory but with a different zero-bias capacitance (e.g. [38,39,42]):

$$C_{j0} = A_E \sqrt{\frac{q}{2V_{Dh}} \left( \frac{\epsilon_A N_A \epsilon_D N_D}{\epsilon_A N_A + \epsilon_D N_D} \right)}. \quad (4.105)$$

In addition, the built-in voltage  $V_{Dh}$  according to (3.152) is different from that of a homojunction since it needs to reflect the spatial variation of the bandgap within the SCR; i.e. the different bandgap at the two ends of the SCR.

In practical applications, the doping is typically strongly non-symmetrical. Assuming  $N_A \gg N_D$ , (4.105) reduces to

$$C_{j0} = A_E \sqrt{\frac{q \epsilon_D N_D}{2V_{Dh}}}. \quad (4.106)$$

As expected, this result is no different from that of a homojunction, except for a different built-in voltage. Usually though, the heterojunction is misaligned by some distance  $\Delta x_h$  from the pn-junction as sketched in Fig. 4.35b. In this case, the voltage dependence and the relations given above change as discussed below.

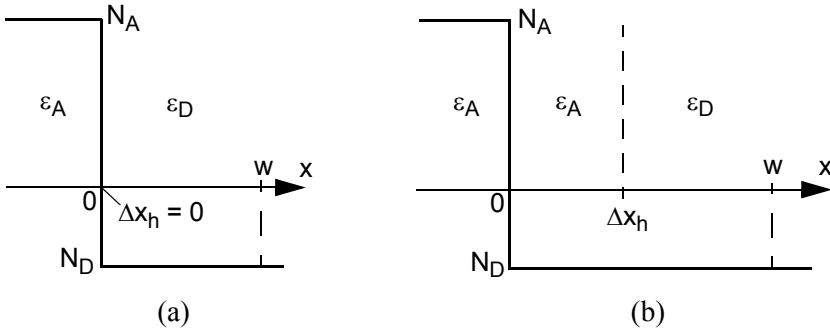


Fig. 4.35: pn-junction with different locations of the heterojunction: (a) alignment of pn- and heterojunction, (b) shift of heterojunction by  $\Delta x_h$  into the donor region.

Since for a strongly non-symmetrical doping the lower doped region contains almost entirely the voltage drop and space-charge region, only the donor side in Fig. 4.35b needs to be considered. As long as the SCR width  $w$  extends beyond  $\Delta x_h$ , Poisson's equation is solved first within the region  $\Delta x_h \leq x \leq w$  and yields after applying the boundary condition  $E(w) = 0$ :

$$E(x) = -\frac{q}{\varepsilon_D} N_D (w - x) \quad \text{for } \Delta x_h \leq x \leq w. \quad (4.107)$$

Then, solving Poisson's equation for the region  $0 \leq x < \Delta x_h$  with the boundary condition of continuous electric displacement at  $\Delta x_h$  gives

$$E(x) = -\frac{q}{\varepsilon_A} N_D (w - x) \quad \text{for } 0 \leq x < \Delta x_h. \quad (4.108)$$

Integrating the field over  $w$  equals the applied effective voltage  $V_D - V$  (neglecting the very small voltage drop in the  $N_A$  region) and leads to a quadratic equation for  $w$ , the solution of which reads

$$w = \Delta x_h \left(1 - \frac{\varepsilon_D}{\varepsilon_A}\right) + \sqrt{\Delta x_h^2 \frac{\varepsilon_D}{\varepsilon_A} \left[\frac{\varepsilon_D}{\varepsilon_A} - 1\right] + \frac{2\varepsilon_D}{qN_D} (V_{Dh} - V)}. \quad (4.109)$$

This reduces to the homojunction form for  $\Delta x_h = 0$  or  $\varepsilon_A = \varepsilon_D$ . On the other hand, if the SCR width is smaller than  $\Delta x_h$  one obtains just the same result as for a homojunction. For the capacitance, given by  $C_j = qA_E N_D (dw/dV)$ , one obtains

$$\frac{C_j}{A_E} = \sqrt{\frac{\varepsilon_D q N_D}{2}} \begin{cases} \frac{1}{\sqrt{\Delta x_h^2 \left[\frac{\varepsilon_D}{\varepsilon_A} - 1\right] \frac{q N_D}{2\varepsilon_A} + (V_{Dh} - V)}}, & w > x_h \\ \frac{\sqrt{\varepsilon_A}}{\sqrt{\varepsilon_D}} \frac{1}{\sqrt{(V_{Dh} - V)}}, & w \leq x_h \end{cases} \quad (4.110)$$

Obviously, two different branches exist for the voltage dependence of the capacitance causing a discontinuity at  $w = \Delta x_h$ , which is the reason for the kink observed in Fig. 4.34. Usually, the shift of the heterojunction is very small, so that the upper relation of (4.110) applies. This can be rewritten again in the compact form (3.139), but now with a different built-in voltage

$$V_{Dhs} = V_{Dh} + \Delta x_h^2 \left[\frac{\varepsilon_D}{\varepsilon_A} - 1\right] \frac{q N_D}{2\varepsilon_A} \quad (4.111)$$

that also needs to be inserted into the zero-bias capacitance (4.106) and (4.105). Depending on the permittivity ratio the new built-in voltage is smaller or larger than the one of a perfectly aligned heterojunction,  $V_{Dh}$ .

For instance, if the p-region in Fig. 4.35a represents the SiGe base of a transistor and the n-region the Si collector, then  $\epsilon_D(\text{Si}) < \epsilon_A(\text{SiGe})$  so that  $V_{Dhs} < V_{Dh}$ . This is the reason for the slightly higher zero-bias value of curve 2 in Fig. 4.34.

Material (and doping) profiles in real transistors are always graded, like in the structure used for Fig. 4.34, since a certain grading cannot be avoided during fabrication. Generally, grading is advantageous as it eliminates transport barriers and discontinuities. However, it usually makes the derivation of closed-form analytical equations much harder. The resulting solution, if possible at all, is often more complicated and does not necessarily provide more insight into the fundamental device physics than the simple equations derived above, which can also easily be applied to other variations of pn- and heterojunctions.

#### 4.4.1.2 Current dependence of base-collector depletion charge

The qualitative analysis of the collector region in, e.g., section 4.2 indicates a strong dependence of the charge  $Q_{BC}$  within the BC SCR on transfer current once the mobile carrier density is no longer negligible compared to the collector doping concentration. Figure 4.36 shows the bias dependence of the associated depletion capacitance  $C_{jCi}$  in the application relevant bias range. The observed behavior is typical for a large variety of investigated doping profiles. According to the theory for  $I_T = 0$  in sec. 3.4, both  $Q_{BC}$  and  $C_{jCi}$  always assume their largest value at forward bias and then decrease monotonously towards the reverse voltage region. At a constant BC voltage, Fig. 4.36 shows that  $C_{jCi}$  decreases rapidly with current beyond the critical current  $I_{CK}$ . The slight decrease at low currents results from the additional negative charge carried by the electrons on the base side of the BC SCR, which adds to the acceptor charge there and requires an increase of the corresponding donor charge on the collector side. This in turn leads to a slight widening of the SCR and, hence, reduction of  $C_{jCi}$  with current. For reverse voltages, the decay with current is monoto-

nous due to the continuous increase of the SCR width with current (cf. (3.32)). At forward bias  $C_{jCi}$  first also decreases slightly but then increases again to values that - for sufficiently high forward bias - can be even larger than the corresponding zero-current value. The strong increase reached before the critical current is caused by the collapse of the BC SCR as described by (3.32).

The current dependence is investigated below quantitatively pursuing two different approaches. The first one is based on simple classical-like considerations while the second one provides a more consistent treatment.

For a spatially constant collector doping profile the charge within the BC SCR is given by

$$Q_{BC} = qA_E(N_{Ci} - n)w_{BC}. \quad (4.112)$$

Assuming a drift current through the SCR, saturation carrier velocity, i.e.  $I_T = qnv_{sn}$ , and inserting (3.32) for  $w_{BC}$  yields

$$Q_{BC} = Q_{BC0} \sqrt{\left(1 - \frac{I_T}{I_{lim}}\right) \left(1 - \frac{V_{B'C} + E_{wc}w_{Ci}}{V_{DCi}}\right)}. \quad (4.113)$$

$I_{lim}$  is defined by (3.4) and the zero-bias charge is given by

$$Q_{BC0} = Q_{BC}(V_{B'C} = 0, V_{B'E} = 0) = qA_EN_{Ci}w_{BC0} \quad (4.114)$$

in which  $w_{BC0}$  represents the SCR width at equilibrium.

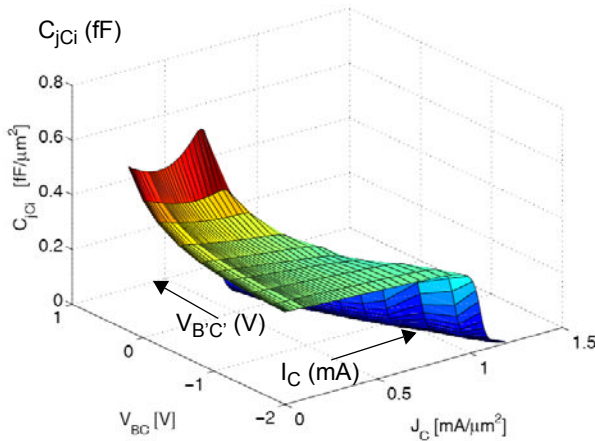


Fig. 4.36: Typical voltage and current dependence of the (area specific) internal BC depletion capacitance for a CED transistor [16].



For small-signal quasi-static operation, the derivatives of  $Q_{BC}$  are of interest. According to (2.49) one can write for the incremental charge

$$dQ_{BC} = C_{cE}dV_{B'E'} + C_{jCi}dV_{B'C} \quad (4.115)$$

with the (internal) BC depletion capacitance (cf. sec. 2.4.1)

$$C_{jCi}(V_{B'C}, V_{B'E'}) = \left. \frac{\partial Q_{BC}}{\partial V_{B'C}} \right|_{V_{B'E'}} \quad (4.116)$$

and the cross-coupling capacitance  $C_{cE}$  defined in (2.44), which represents the mobile charge. Since  $Q_{BC}$  according to (4.113) is an explicit function of current (rather than  $V_{B'E'}$ ), often a different selection of the independent variables is used to calculate the incremental charge:

$$dQ_{BC} = \tau_{BC}^* dI_T + C_{jCi}^* dV_{B'C}. \quad (4.117)$$

In this case, the capacitance is defined somewhat differently as

$$C_{jCi}^*(V_{B'C}, I_T) = \left. \frac{\partial Q_{BC}}{\partial V_{B'C}} \right|_{I_T}. \quad (4.118)$$

By keeping a constant current instead of a constant  $V_{B'E'}$ , (as in (4.116)), the impact of the Early-effect and corresponding mobile charge on the value of the capacitance at higher current densities is significantly reduced in (4.118); in other words, (4.118) only contains the impact of the carrier velocity change on the depletion capacitance value. From a compact modeling point of view it is more convenient to use (4.117) and (4.118), while device simulation usually calculates the capacitance from (4.116).

Employing (4.118) and the charge from (4.113), the current dependent depletion capacitance reads

$$C_{jCi}^*(V_{B'C}, I_T) = C_{jCi0} \sqrt{1 - \frac{I_T}{I_{lim}}} \frac{1 + w_{Ci} \left. \frac{dE_{wc}}{dV_{B'C}} \right|_{I_T}}{\sqrt{1 - \frac{V_{B'C} + E_{wc} w_{Ci}}{V_{DCi}}}}. \quad (4.119)$$

The term  $\sqrt{1 - I_T/I_{lim}}$  in the numerator represents the widening of the SCR at larger (reverse) voltages, while the  $E_{wc}$  term in the denominator

represents the collapse of the SCR towards forward bias. Obviously, the result still depends on the expression inserted for the field  $E_{wc}$ . If the latter is described by just an ohmic voltage drop with a linear current dependence the corresponding derivative is zero. For a compact model, the equation also needs to be made numerically stable, e.g., by suitably smoothing the square-root arguments in the charge expression (4.113). In addition, the exponent  $1/2$  (i.e. the square-root) in (4.113) can be replaced by the variable  $z$  as in the classical theory of sec. 3.4.1 in order to make the expression applicable to a wider range of junction structures.

Figure 4.37 compares at selected voltages  $V_{B'C'}$ , the current dependence of  $C_{jCi}$  obtained from device simulation results according to (4.116) with the analytical relation (4.119), in which  $E_{wc}$  is described by (4.18). The analytical results agree quite well at lower currents up to approximately the current at which arithmetic overflow would occur and the curve was truncated. The deviation in the voltage dependence at low currents can be eliminated by using an exponent coefficient  $z$  that is different from  $1/2$ .

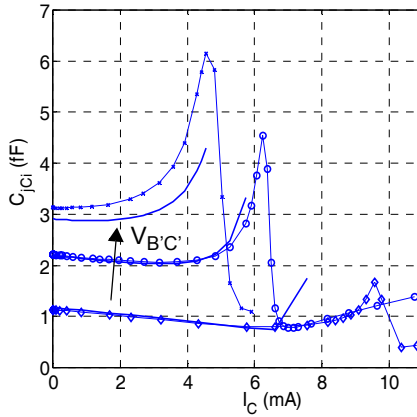


Fig. 4.37: Current dependence of the base-collector depletion capacitance for different voltages  $V_{B'C'}/V = 0.4, 0, -2.3$ . Comparison between device simulation (thin lines with symbols) and analytical equation (thick solid lines) for the CED-HBT.

$Q_{BC}$  of (4.113) does not equal  $Q_{jCi}$  of sec. 3.4 since the latter is defined with respect to  $V_{B'C'} = 0$  as reference. Hence,  $Q_{jCi}(V_{B'C'}=0) = 0$  and

$$Q_{jCi} = Q_{BC0} - Q_{BC}. \quad (4.120)$$

Inserting (4.113) with the exponent parameter  $z$  yields for zero current the same result as (3.141). Note, that the punch-through case is not included in the results above and would need to be added.

An approach that is more general and also consistent with the already existing analysis of the mobile charge related time constant  $\tau_{BC}$  is based on the relation

$$Q_{BC} = \int_{x_{mc}}^{x_{mc} + w_{Ci}} \rho(\xi) d\xi = \varepsilon(E_{jc} - E_{wc}). \quad (4.121)$$

Here, the definitions  $E_{jc} = -E(x_{jc})$  and  $E_{wc} = -E(w_{jci})$  of sec. 3.2 were used. The lower integration limit is given by the carrier cross-over point  $x_{mc}$  (cf. (2.28)), and the integral represents the total collector space charge under all bias conditions. According to (4.121), the internal BC depletion capacitance can now be calculated as a function of voltage *and current* from the formulation of the electric field.

In order to link the electric field with the measurable charge  $Q_{jCi}$ , an equilibrium (reference) field is introduced,

$$E_{jc00} = E_{jc}(V_{B'C} = 0, I_T = 0) = \frac{Q_{BC0}}{\varepsilon}, \quad (4.122)$$

assuming  $E_{wc00} = 0$ , which is reasonable for practical structures. Inserting this together with (4.121) into (4.120) yields

$$E_{jc} = E_{jc00} - \frac{Q_{jCi}}{\varepsilon} + E_{wc} \quad (4.123)$$

or in normalized form

$$\frac{E_{jc}}{E_{jc00}} = 1 - \frac{Q_{jCi}}{\varepsilon E_{jc00}} + \frac{E_{wc}}{E_{jc00}}. \quad (4.124)$$

This relation can be used to calibrate the field model at low current densities where  $E_{wc}$  is negligible. The exact expression for the capacitance and BC time constant depend on the selected (and smoothed) description for the field components.

#### 4.4.2 Mobile charge

In this section the impact of the material composition on the mobile carrier related charge in the various transistor regions is discussed. In contrast to sec. 3.5, each charge and corresponding transit time component is now considered separately over the whole bias region. The analytical equations of sec. 3.5 serve as a basis, which is extended appropriately where required.

Figure 4.38 provides an overview on the relevance of the various storage times for two different transistor versions. As for BJTs, in both cases the BE SCR storage time dominates at very low current densities. For the high-speed HBT version (Fig. 4.38a) the base transit time dominates over almost the entire practically relevant bias region, closely followed by the delay time  $\tau_{BC}$  through the BC SCR. In contrast, for the power transistor version (Fig. 4.38b)  $\tau_{BC}$  is as important as the neutral base component. At high current densities, the BC barrier effect causes in both cases a rapid increase of all components, led by the base storage time and followed by the collector storage time. In all cases, the emitter storage time plays only a minor role.

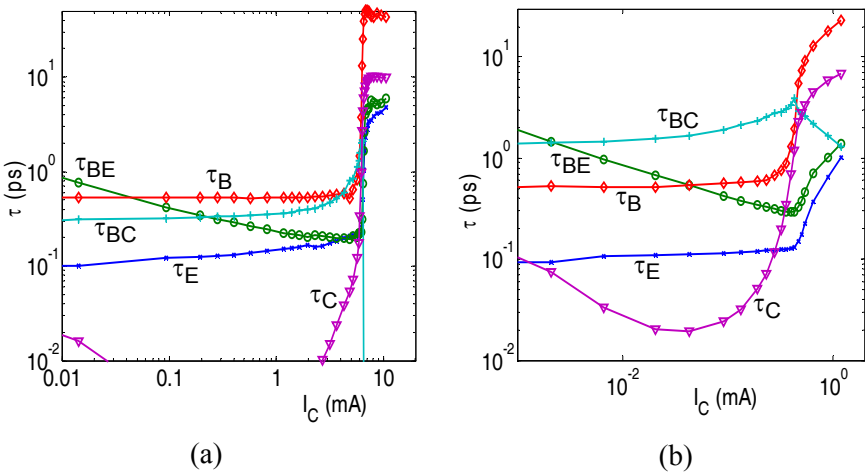


Fig. 4.38: Storage times (log) calculated from device simulation using the Regional Approach vs. collector current (log) at  $V_{B,C} = 0$  for two different transistor versions of the CED HBT: (a)  $N_{Ci} = 5 \cdot 10^{17} \text{ cm}^{-3}$ ; (b)  $N_{Ci} = 4 \cdot 10^{16} \text{ cm}^{-3}$ .

#### 4.4.2.1 Neutral base region

The impact of a variable material composition on minority charge and transit time can be analytically evaluated following a theory outlined in [21]. The derivation starts with applying the electron transport equation (4.28) to the neutral base region. Realizing that the hole QFP,  $\phi_p$ , is spatially constant in the neutral base, extension of (4.28) by  $d\phi_p/dx = 0$  and replacing the electron current density by the transfer current yields

$$I_T = -qA_E\mu_n n \frac{d(\phi_p - \phi_n)}{dx}. \quad (4.125)$$

The QFP difference is related to the pn-product using (2.4a) and (2.4b),

$$pn = n_i^2 \exp\left(\frac{\phi_p - \phi_n}{V_T}\right), \quad (4.126)$$

with the effective intrinsic carrier density from (2.7). Taking the derivative

$$\frac{d(\phi_p - \phi_n)}{dx} = V_T \frac{n_i^2}{pn} \frac{d(pn/n_i^2)}{dx}, \quad (4.127)$$

isolating it in (4.125) and replacing it then by the above expression gives after integration from some point  $x'$  in the neutral base to its end  $w_B$

$$\int_{x'}^{w_B} \frac{-I_T}{qA_E\mu_n V_T n_i^2} p dx' = \left. \frac{pn}{n_i^2} \right|_{w_B} - \left. \frac{pn}{n_i^2} \right|_{x'}. \quad (4.128)$$

For a sufficiently reverse biased BC junction the first term on the r.h.s. is negligible. Bringing  $p/n_i^2$  of the remaining term to the l.h.s. allows to write a generalized expression for the minority charge in the neutral base:

$$Q_{nBf} = qA_E \int_0^{w_B} n dx' = \frac{I_T}{V_T} \int_0^{w_B} \frac{n_i^2}{p} \int_{x'}^{w_B} \frac{1}{\mu_n n_i^2} p d\xi dx'. \quad (4.129)$$

Assuming an average mobility  $\bar{\mu}_{nB}$  and low current densities (i.e. negligible variation of the base width with current), one obtains for the low-current base transit time

$$\tau_{nBf0} = \frac{Q_{nBf}}{I_T} = \frac{1}{\bar{\mu}_{nB} V_T} \int_0^{w_B} \frac{n_i^2}{p} \int_{x'}^{w_B} \frac{p}{n_i^2} d\xi dx'. \quad (4.130)$$

This result is a generalized form of the well-known Moll-Ross equation [23] for the transit time. It allows to evaluate the impact of bandgap variations on transit time (at low injection) if a specific bandgap profile is inserted.

For illustration purposes, a trapezoid Ge profile (cf. curve 2 in Fig. 4.6) shall be assumed, in which the bandgap varies linearly across the neutral base region. Hence, the composition dependent energy term reads

$$\Delta W_m = -qE_m x' \quad (4.131)$$

with the drift field  $E_m$  ( $< 0$ ) defined by the composition variation. For the sake of simplicity, a spatially constant doping profile is assumed, so that  $p = N_B$  and also bandgap variations due to high-doping effects can be neglected. Therefore, inserting (4.131) into (2.6) yields for constant (reference) temperature

$$V_n + V_p = V_T \ln \left( \frac{N_C}{N_{C, Si}} \frac{N_V}{N_{V, Si}} \right) - E_m x'. \quad (4.132)$$

Lumping the relatively weak composition dependence of the density of states and of the mobility into spatial average values gives with (2.7) for (4.130) after evaluation

$$\tau_{nBf0} = \frac{w_B^2}{\bar{\mu}_{nB} V_T \zeta_m} \left[ 1 - \frac{1 - \exp(-\zeta_m)}{\zeta_m} \right]. \quad (4.133)$$

with the material composition dependent drift factor

$$\zeta_m = -\frac{E_m w_B}{V_T} > 0. \quad (4.134)$$

For a large drift field the term in the bracket approaches 1, so that  $\tau_{nBf0} \sim 1/\zeta_m$ .

Defining, as in sec. 3.3.1.1, a drift function  $f_z = \exp(\zeta_m)$ , and using  $F_z$  from (3.172) gives

$$\tau_{nBf0} = \frac{w_B^2}{F_\zeta \bar{\mu}_{nB} V_T} \quad (4.135)$$

which is identical with  $\tau_{Bfd}$  from (3.171). This result is expected since in the drift-transistor theory of sec. 3.3 and 3.5 no assumptions about the origin of the electric field in the base were made. As a consequence, also the case of a spatially constant material composition in the base can be deduced from either (4.133) or the expression in sec. 3.5 yielding the familiar result (3.173). In other words, a Ge box profile with no grading does not yield any improvement in the base transit time.

For medium and high current densities, the calculation of the electron density and charge in the base proceeds the same way as described in sec. 3.5, except that the boundary condition for  $n_c = n(x_c)$  is different if a material composition variation exists in the BC SCR. Starting point is the differential equation (3.44), in which  $\bar{E}_n$  contains also the field contribution from intentional material composition. The solution is given by (3.45) and, after integration, yields the charge

$$Q_{nB} = \frac{w_B^2(I_T)}{F_\zeta \bar{\mu}_{nB} V_T} I_T + q A_E \frac{w_B(I_T)}{G_\zeta} n_c(I_T) \quad (4.136)$$

with  $F_z$  and  $G_z$ , respectively, from (3.172) and (3.175), respectively.

Neglecting the material dependence of the electron affinity and density of states, the electron density  $n_c$  at low and medium current densities is given by (3.58) until the barrier  $\Delta W_{Cb}$  in the conduction band starts forming. As discussed in section 4.1.3.3, the barrier leads also to the drop  $\Delta V_{Cb} = \Delta W_{Cb}/(-q)$  in the electrostatic potential. With the neglects made above, the ratio of the electron density  $n_c$  at the end of the neutral base (that corresponds to the start of the dipole layer) and the density  $n_{jc}$  at the beginning of the collector (that corresponds to the end of the dipole layer) is given by

$$\frac{n_c}{n_{jc}} = \exp\left(\frac{\Delta V_{Cb}}{V_T}\right). \quad (4.137)$$

For this, the electron quasi-Fermi level has been assumed to be spatially constant across the dipole layer due to the high current densities required

for triggering the barrier formation. This is a reasonable assumption although it is not completely fulfilled according to device simulation. Within the current range, in which the barrier layer starts forming,  $n_{jc}$  is still given by (3.58), so that a rough approximation of  $n_c$  is

$$n_c = \frac{I_T}{A_E q v_c} \exp\left(\frac{\Delta V_{Cb}}{V_T}\right). \quad (4.138)$$

Inserting this into (4.136) yields

$$Q_{nB} = \left[ \frac{w_B^2(I_T)}{F_\zeta \bar{\mu}_{nB} V_T} + \frac{w_B(I_T)}{G_\zeta v_c(I_T)} \exp\left(\frac{\Delta V_{Cb}(I_T)}{V_T}\right) \right] I_T, \quad (4.139)$$

where “( $I_T$ )” indicates the major current dependent variables. The derivative w.r.t. current gives the corresponding components of the base transit time. While the drift/diffusion component  $\tau_{Bfd}$  does not change, the carrier jam component now looks different from (3.242):

$$\tau_{Bfvh} = \left[ \tau_{Bfv} + \frac{w_B}{G_\zeta v_c} \frac{I_T}{V_T} \frac{d\Delta V_{Cb}}{dI_T} \bigg|_{V_{CE}} \right] \exp\left(\frac{\Delta V_{Cb}}{V_T}\right). \quad (4.140)$$

Here, the index “h” has been added to distinguish the result from the homojunction transit time  $\tau_{Bfv}$ . There are two new terms in the expression above. First, a term containing the derivative of the current dependent barrier is added to  $\tau_{Bfv}$ . This term starts to contribute at  $I_{CK}$ . Second, this sum is multiplied with the exponential dependence on the barrier potential. Since the latter remains zero up to  $I_{CK}$ ,  $\tau_{Bfvh} = \tau_{Bfv}$  at low and medium current densities. However,  $\tau_{Bfvh}$  is expected to increase very rapidly beyond  $I_{CK}$ . Due to the exponential dependence, a description is needed for  $\Delta V_{Cb}$  in a compact model that can be flexibly adapted to measured data.

#### 4.4.2.2 Neutral collector

As discussed in section 4.1.3.3 minority charge storage in the collector starts *after* the barrier region has formed. This is equivalent to a sufficient reduction in the valence band barrier for holes, permitting holes to be injected into the collector. This in turn leads to the formation of an injection



zone in a very similar way as discussed in sec. 3.5.3.1 for BJTs. Therefore, the same theory can be applied. Notice though that when the injection zone starts to form in a SiGe HBT the base storage time has already increased significantly due to the barrier effect in contrast to a BJT in which the significant increase in base storage time starts with the formation of the injection zone.

The expressions (3.272) and (3.273), respectively, for the stored excess collector hole charge and storage time, respectively, depend on the injection width  $w_i$ . Depending on the collector voltage  $V_{ci}$  two different solutions are obtained for  $w_i$ , namely (3.261) for low voltages and (3.265) for high voltages. Figure 4.39 shows the injection width of the HBT as a function of current for different voltages. The results obtained from device simulation are compared to the analytical expressions (3.261) and (3.265). The high-voltage expression does not start to increase below  $I_{lim}$ , and even with already reduced values of the critical current (due to  $V_{lim} > 0$  in  $I_{CK}$  compared to the literature) the increase at the highest voltage starts too late. Apparently, as for the BJT in Fig. 3.25, the accuracy of the high-voltage expression (3.265) seems to be worse than the accuracy of the low-voltage expression (3.261), especially at lower voltages. This is in contrast to the claims made in [45], which were not corroborated though by device simulation results and only discussed for higher voltages. Furthermore, while the low-voltage expression is numerically applicable over the entire voltage range, the high-voltage expression requires significant numerical conditioning to be applicable to the low-voltage region. As a consequence, the low-voltage expression (3.261) appears to be more suitable for compact modeling.

Although the injection width from (3.261) predicts the *onset* of the increase fairly well, the slope of the curve is too small. By a very simple extension of the low-voltage expression (3.261) an excellent approximation of  $w_i$  over the entire bias range can be obtained. Introducing an exponent factor  $\beta_{wi}$  in the argument of the normalized injection width,

$$\frac{w_i}{w_{Ci}} = 1 - \left( \frac{I_{CK}}{I_T} \right)^{\beta_{wi}}, \quad (4.141)$$

yields the dashed lines in Fig. 4.39, which are in excellent agreement with the device simulation results, especially for lower voltages. In this demon-

stration, the critical current calculated from process parameters also was decreased slightly, and the smoothing function for  $w_i/w_{Ci}$  used in HICUM [46] (see also ch. 8) was applied to (4.141). The parameters were adjusted based on the storage time behavior shown further below.

The collector charge and storage time, respectively, are still given by the expressions (3.272) and (3.273), respectively, if the charge in the very narrow barrier region is neglected. The latter assumption is a reasonable approximation at not too high currents, i.e. just beyond  $I_{CK}$ , which is also the most important region from a practical application point of view. Inserting the injection width from device simulation into (3.273) yields the storage time shown in Fig. 4.40. All parameters were calculated directly from the device structure. Note that the high-voltage expression of  $I_{CK}$  as defined in (3.250) is already lower (due to the subtraction of  $V_{lim}$ ) than the corresponding expressions still used in the more recent literature (e.g. [11,12,13,45]). In other words, using just (3.250) with  $V_{lim} = 0$  yields by far too large values for the onset of high-current effects at all voltages and, thus, is unsuitable for compact modeling!

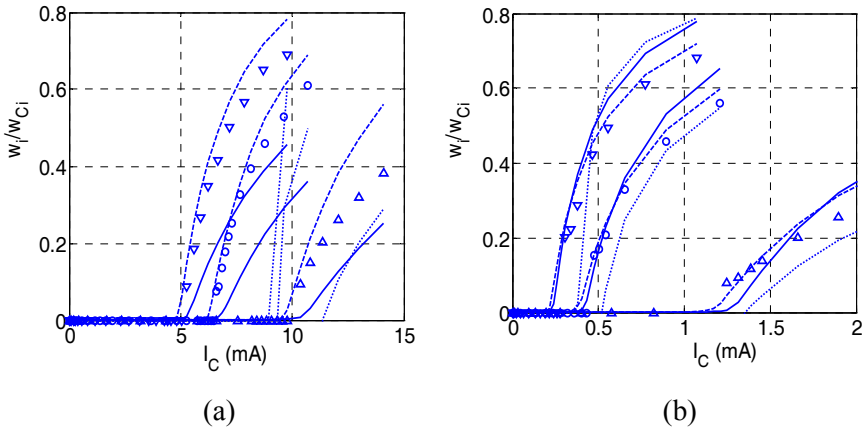


Fig. 4.39: Normalized injection zone width as function of collector (transfer) current for two different versions of the 1D CED HBT: (a) High-speed version with  $N_{Ci} = 5 \cdot 10^{17} \text{cm}^{-3}$  and (b) high-voltage version with  $N_{Ci} = 2 \cdot 10^{16} \text{cm}^{-3}$ . Comparison of the analytical expressions (3.261) (solid lines) and (3.265) (dotted lines), calculated from process parameters, with device simulation results obtained using REGAP (symbols). In addition, the modified version (4.141) has been inserted with (a)  $\beta_{wi} = 2.2$ , 8%  $I_{CK}$  reduction; (b)  $\beta_{wi} = 0.78$ , 5%  $I_{CK}$  reduction.

The lower saturation value of the analytical model at very high currents in Fig. 4.40 is assumed to result from neglecting the charge in the barrier region. In a compact model, this discrepancy as well as the somewhat larger critical current can be easily eliminated by properly adjusting the saturation storage time  $\tau_{pCs}$  from (3.274) and reducing  $I_{CK}$  by increasing  $r_{Ci0}$ . If in addition (4.141) is used for  $w_i$  excellent agreement can be obtained as shown by the dashed lines in Fig. 4.40. It is interesting to note that the analytical model seems to be more accurate for transistors with lower collector doping.

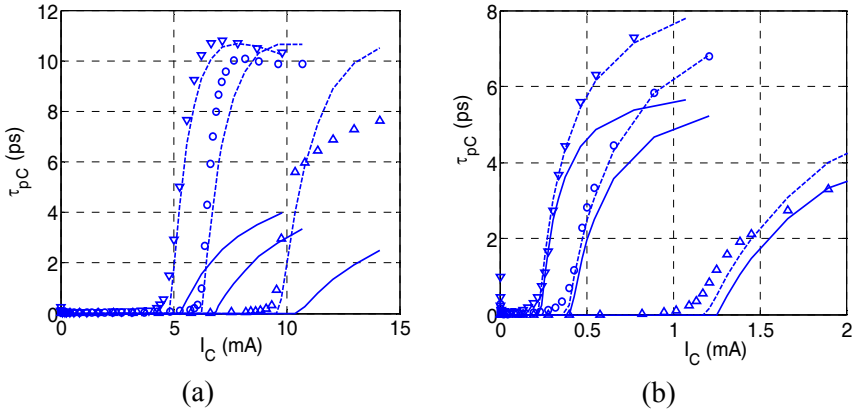


Fig. 4.40: Collector storage time as function of collector (transfer) current for two different versions of the 1D CED HBT: (a)  $N_{Ci} = 5 \cdot 10^{17} \text{ cm}^{-3}$ ; (b)  $N_{Ci} = 2 \cdot 10^{16} \text{ cm}^{-3}$ . Comparison of the analytical expression (3.273) (dashed lines) with device simulation results obtained using REGAP (symbols). In addition, (3.273) was evaluated with  $w_i$  from (4.141) (dashed lines). Parameters: (a)  $\beta_{wi} = 2.2$ ,  $\tau_{pCs} = 7.5 \text{ ps}$ , 8%  $I_{CK}$  reduction; (b)  $\beta_{wi} = 0.78$ ,  $\tau_{pCs} = 9.5 \text{ ps}$ , 5%  $I_{CK}$  reduction.

In practice, the heterojunction is often shifted somewhat into the collector region. In this case, the injection region will initially start to form the same way as in a BJT until it reaches the heterojunction. At this point, the holes will pile up at the heterojunction due to the valence band barrier, and the dipole layer again starts to form. From this point on the same mechanisms occur as discussed before for the case of the heterojunction being aligned with the BC junction. If the distance between the BC junction and the heterojunction is large enough, one can observe two distinct shapes of the overall storage time (or  $f_T$ ) [33]. However, the current densities at

which this difference is visible is generally far outside the practical application range for the transistor.

#### 4.4.2.3 Base-collector depletion region

At low current densities the BC heterojunction behaves electrically the same way as a homojunction: electrons reaching the end of the neutral base are swept through the BC SCR as a drift current. The exact velocity depends on the voltage  $V_{C'E'} > 0$  or the corresponding value of  $V_{ci}$  (cf. Figs. in section 4.2). Therefore, the same considerations and theoretical results that were obtained for  $\tau_{BC}$  in sec. 3.5 also apply for HBTs.

Once the dipole layer and barrier start to form, the resulting retarding field causes a significant gradient of the electron density to drive the transfer current through the barrier region by diffusion (cf. Figs. 4.8 and 4.9). An analysis of the excess charge stored in the barrier region yields as a rough estimate for the upper limit of the corresponding storage time

$$\tau_{Cb} \approx k_{Cb} \frac{w_{Cb}}{v_{sn}} \quad (4.142)$$

with  $k_{Cb} \approx 2 \dots 5$ . Here, it was assumed that the barrier region  $w_{Cb}$  has completely formed and that the injection region does not yet exist. Although in this case the factor  $k_{Cb}$  is larger than 1, the value of  $\tau_{Cb}$  is generally very small compared to  $\tau_{BC}$  and can be included in  $\tau_{BC}$  due to its similar bias dependence and the fact that it cannot be measured directly.

#### 4.4.2.4 Neutral emitter

The minority charge in the neutral emitter region depends on the location of a possible heterojunction in the BE region. Assuming again a trapezoid Ge profile with a step that causes a small bandgap difference somewhere in the BE region (cf. curve 2 in Fig. 4.6), various cases can be distinguished. As long as the heterojunction is located in the BE SCR it has, according to section 4.3.2, negligible impact on the backinjection current and associated charge for a given value of  $V_{B'E'}$ . Nevertheless, the emitter storage time is still reduced due to the transfer current and, hence,

current gain enhancement caused by the addition of Ge to the base region, i.e.

$$\frac{\tau_{Ef, SiGe}}{\tau_{Ef, Si}} = \frac{\beta_{f0, Si}}{\beta_{f0, SiGe}}. \quad (4.143)$$

If the heterojunction is located within the neutral emitter, as shown in Fig. 4.28, the situation changes. Obviously, the charge will increase in the SiGe portion  $x \in [x_{he}, w_E]$  due to the increased intrinsic carrier density there. In other words, at the same  $V_{B'E}$ , the backinjection carrier density at  $w_E$  will now correspond to that in SiGe material instead of Si. Revisiting the calculations performed in sec. 3.5.1.3 yields for the SiGe portion the carrier density

$$\Delta p_h(x) = \frac{I_{jBEi}(w_E - x_{he})}{qA_E \bar{\mu}_{pEh} V_T} \left( \frac{x - x_{he}}{w_E - x_{he}} \right) + \Delta p_h(x_{he}). \quad (4.144)$$

The “offset” density  $\Delta p_h(x_{he})$  at the heterojunction is related through (4.90) to the density  $\Delta p_m(x_{he})$  in the Si region. The latter in turn is given by (3.209) if  $w_E$  is replaced by  $x_{he}$ . The carrier density at the poly-mono interface remains the same as (3.214). Integrating (4.144) yields for the minority charge in the SiGe region

$$Q_{pEh} = \frac{I_{jBEi}}{2\bar{\mu}_{pEh} V_T} (w_E - x_{he})^2 + qA_E \Delta p_h(x_{he})(w_E - x_{he}). \quad (4.145)$$

The charge  $Q_{pEm}$  in the mono-Si region is given by (3.215) if  $w_E$  is replaced by  $x_{he}$ . The total minority emitter charge now is the sum

$$Q_{pE} = Q_{pEh} + Q_{pEm} + Q_{pEp}.$$

The corresponding storage time  $\tau_{Efp}$  in the poly-region is the same as (3.219), while the storage time  $\tau_{Efm}$  is given by (3.218) if  $w_E$  is replaced by  $x_{he}$ . Finally, the additional contribution from the SiGe region is calculated from (4.145) as

$$\tau_{Efh} = \frac{1}{\beta_{f0}} \left( \frac{(w_E - x_{he})^2}{2\bar{\mu}_{pEh} V_T} + \left[ \frac{w_E - x_{he}}{\bar{\mu}_{pEm} V_T} x_{he} + \frac{w_E - x_{he}}{\frac{\bar{\mu}_{pEp} V_T}{L_{pp}} \coth\left(\frac{w_p}{L_{pp}}\right) + v_{pmI}} \right] \exp\left(\frac{\Delta V_V}{V_T}\right) \right).$$

The first term corresponds to the diffusion of carriers towards the heterojunction. The last term represents the “offset” charge caused by the finite diffusion velocity in the mono- and poly-Si region, which is increased by the valence band barrier term to a value that corresponds to the smaller bandgap in the SiGe region. Obviously,  $\tau_{Efh}$  drops to zero if the heterojunction location approaches  $w_E$ .

The original expressions for  $\tau_{Efh}$  as well as for  $\tau_{Efm}$  and  $\tau_{Efp}$  are fairly lengthy and difficult to overview as a sum. Often, as already mentioned in sec. 3.5.1.3, the interface recombination velocity  $v_{pmI}$  can be assumed to be large compared to the  $L_{pp}$  term in the denominator, so that one obtains for the total emitter storage time

$$\tau_{Ef} = \frac{1}{\beta_{f0}} \left( \frac{(w_E - x_{he})^2}{2\bar{\mu}_{pEh} V_T} + \frac{x_{he}^2}{2\bar{\mu}_{pEm} V_T} + \frac{w_E - x_{he}}{\bar{\mu}_{pEm} V_T} x_{he} \exp\left(\frac{\Delta V_V}{V_T}\right) + \frac{(w_E - x_{he}) \exp\left(\frac{\Delta V_V}{V_T}\right) + x_{he} + L_{pp} \frac{\cosh(w_p/L_{pp}) - 1}{\sinh(w_p/L_{pp})}}{v_{pmI}} \right).$$

Assuming  $w_E - x_{he} < x_{he}$ ,  $\bar{\mu}_{pEh} \approx \bar{\mu}_{pEm}$ , and  $\exp(\Delta V_V/V_T) > 1$  as well as an effective interface recombination velocity

$$v_{Epm} = \frac{v_{pmI}}{1 + \frac{w_E - x_{he}}{x_{he}} \exp\left(\frac{\Delta V_V}{V_T}\right) + \frac{L_{pp}}{x_{he}} \frac{\cosh(w_p/L_{pp}) - 1}{\sinh(w_p/L_{pp})}}, \quad (4.146)$$

the storage time reduces to a more compact expression

$$\tau_{Ef} = \frac{1}{\beta_{f0}} \left( \frac{x_{he}^2}{2\bar{\mu}_{pEm} V_T} + \frac{w_E - x_{he}}{\bar{\mu}_{pEm} V_T} x_{he} \exp\left(\frac{\Delta V_V}{V_T}\right) + \frac{x_{he}}{v_{pm}} \right). \quad (4.147)$$

Here, the SiGe contribution is roughly represented by the middle term, while the first term corresponds to the mono-Si region and the last term to the poly-Si region. Note, that the emitter charge and storage time are usually negligible in SiGe HBTs due to the valence band barrier in trapezoid Ge profiles and due to the generally higher current gain.

#### 4.4.2.5 Base-emitter depletion region

The carrier distribution within the BE SCR depends on the location of the heterojunction  $x_h$ . As long as  $x_h$  does not fall into the BE SCR the considerations and results of sec. 3.5.1.4 apply, but possibly with different material parameters. Specifically, for  $x_h < x_e$  in Fig. 4.41 SiGe material parameters have to be inserted. As a consequence, the resulting storage time  $\tau_{BE}$  is larger than in BJTs for a given  $V_{B'E'}$  (due to the higher intrinsic carrier density in SiGe). On the other hand, for  $x_h > x_e$  the same Si parameters as in sec. 3.5.1.4 have to be inserted. Now,  $\tau_{BE}$  is the same as in BJTs for given  $V_{B'E'}$ .

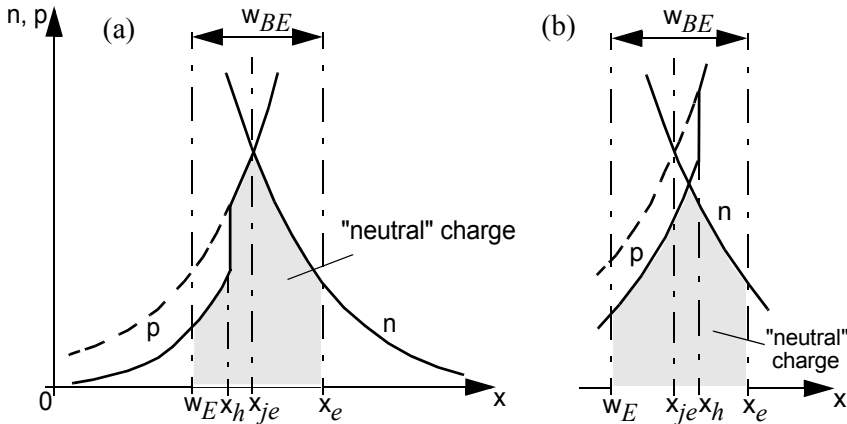


Fig. 4.41: Sketch of carrier densities in the BE SCR for different locations  $x_h$  of the heterojunction within the SCR  $w_{BE}$ .

The situation becomes a bit more complicated for those cases in which  $x_h$  falls into the BE SCR as sketched in Fig. 4.41. Assuming an abrupt material variation at  $x_h$ , the hole density drops abruptly from the SiGe side to the Si side. As a consequence, the charge in the Si region  $x \in [w_E, x_h]$  is smaller compared to that of the homojunction case in sec. 3.5.1.4, which is indicated in Fig. 4.41 by the dashed lines for the hole density. A detailed theory taking into account the different material regions results in lengthy expressions, but still produces the same exponential  $V_{B'E'}$  dependence for the charge and storage time as in sec. 3.5.1.4. This is also confirmed in Fig. 4.38 which shows a similar behavior for  $\tau_{BE}$  as Fig. 3.20 for BJTs.

## 4.5 Vertical non-quasi-static effects

### 4.5.1 Introduction

The considerations so far are based on the assumption that the carrier densities in a vertical (1D) structure follow *instantaneously* the voltage applied across a junction. This operation is called quasi-static (QS), and the associated carrier profiles then lead to the QS charge and current. However, at sufficiently high frequencies or temporal change the finite transit time of carriers through the transistor regions causes a “delayed” reaction with respect to the applied voltage for those carriers that are further away from the junction. This so-called non-quasi-static (NQS) behavior applies in particular to minority carriers in neutral regions while carriers in the vicinity of the depletion region edges would experience no significant delay. Therefore, depletion charges and their associated currents can be treated QS up to very high frequencies, while vertical NQS effects need to be taken into account for minority charges and associated currents as well as transport through (wider) SCRs.

For a given transistor, vertical NQS effects can be observed in the time or frequency dependence of the collector and base terminal current. As an illustrative example, the CED transistor is biased at peak  $f_T$ , and a sinusoidal signal with 1mV amplitude and 100GHz frequency is applied to the base. The resulting time dependence of various currents along with the controlling voltage  $v_{B'E'}(t)$  as reference are shown in Fig. 4.42. At the start



of the transient, the collector current remains unchanged at its DC value for a certain time and then starts to follow  $v_{B'E}(t)$ . On the other hand, the base current jumps at the beginning to a large value and then leads  $v_{B'E}(t)$ . This jump is caused by charging mainly the depletion capacitances.

As can be observed from Fig. 4.42, the collector current keeps lagging  $v_{B'E}(t)$  by a certain delay time  $\tau_{IT}$ . It is possible (see later) to calculate from device simulated carrier distributions also the QS transfer current  $i_{T,QS}$ . As expected, it follows  $v_{B'E}(t)$  directly and has a larger amplitude than the collector current. Note, that in this example current associated with the internal BC depletion capacitance has negligible impact on the delay and magnitude of  $i_C$ . Thus, a QS compact model exhibits a significant error in modeling the time and frequency dependent collector current.

Since the frequency (i.e. temporal change  $\partial/\partial t$ ) is very high, the base current is mostly a charging current and its QS value is expected to lead  $v_{B'E}(t)$  by  $90^\circ$ . However, the actual phase shift, taken at the zero-crossing within the second period (cf. Fig. 4.42) turns out to be less than  $90^\circ$ . In other words, the base current is lagging behind its QS value by a certain delay time  $\tau_{Qf}$ .

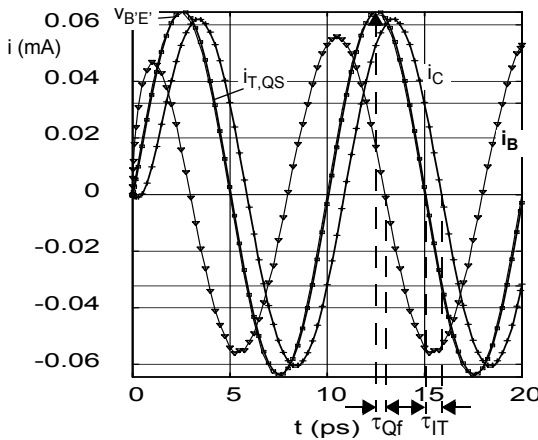


Fig. 4.42: Time dependent base current  $i_B$  ( $\nabla$ ) and collector current  $i_C$  (+) in response to a 100 GHz sinusoidal signal voltage  $v_{B'E}$  (solid line) with 1mV amplitude. In addition, the QS transfer current (o), numerically calculated from GICCR, as well as the numerically calculated collector current from TICCR (solid line through plus signs) have been inserted.

Both time delays  $\tau_{IT}$  and  $\tau_{Qf}$  are caused by the delayed reaction of the minority carriers w.r.t.  $v_{B'E}(t)$ . In practice, measurements of high-speed transients in time domain are very difficult. Hence, the most simple detection of NQS effects is through a small-signal measurements in frequency domain using a network analyzer. The additional phase shifts can then be observed in the input admittance and the transconductance.

The observation of the additional phase shift, represented by characteristic time constants, has led to the often used and heuristic “delay time approach”, in which the quasi-static transconductance is multiplied with the ideal delay term  $\exp(-j\omega\tau_{IT})$  (e.g. [47, 48, 49]). While this approach works for analytical calculations it is not suitable for circuit simulation since an ideal delay term cannot be realized in a time-domain implementation. The goal for a compact model is to include a description that is consistent between time and frequency domain as well as physics-based and computationally efficient. Obviously, the solution approach has to involve or at least include the continuity equation with its time derivative.

In the following sections the main existing theoretical solutions for describing NQS effects are presented and discussed from a physics-based point of view. In order to obtain tractable analytical solutions certain simplifications are required. As a consequence, most solutions are not directly suitable for compact modeling, but can be used as guidelines. Depending on the simplifications the resulting expressions lend itself to a variety of interpretations in terms of a equivalent circuits or other types of model implementation that can be found in literature. The issues and merits of these different implementations will be discussed in detail in ch. 8.

#### 4.5.2 Time-domain description

NQS effects were observed first during fast switching and high-frequency operation of pn diodes. Corresponding theoretical investigations can be found in, e.g. [50], and were based on the exact solution of the continuity equation in a neutral region. The approach and solution can be easily transferred to the base region of bipolar transistors.

Assuming a 1D npn transistor and negligible recombination, the time dependent electron density within the neutral base region  $x' \in [0, w_B]$  is giv-

en by combining the continuity equation (2.8b) and the transport equation (2.9b):

$$\frac{\partial n}{\partial t} = \bar{\mu}_{nB} V_T \frac{d^2 n}{dx'^2} + \bar{\mu}_{nB} E_n \frac{dn}{dx'}. \quad (4.148)$$

Here, the mobility and effective electric field have already been defined as spatially averaged values, which can still be bias dependent though. In order to obtain analytical solutions and a basic understanding of the NQS effect the electric field is neglected and low injection is assumed. This reduces (4.148) to a simple diffusion equation,

$$\frac{\partial n}{\partial t} = D_n \frac{d^2 n}{dx'^2}, \quad (4.149)$$

with  $D_n = \bar{\mu}_{nB} V_T$  as the diffusion constant. The boundary conditions at low injection are

$$n(0, t) = n_e(t) = n_{e0} \exp\left(\frac{v_B E(t)}{V_T}\right), \quad n(w_B, t) = n_{c0} \approx 0. \quad (4.150)$$

Generally, the electron density can be partitioned into a QS and a NQS component. More specifically, one can define a “family” time  $t' = \alpha t$  and write  $n(x', t')$  as a series expansion (e.g. [51])

$$n(x', t') = n_0(x', t') + n_1(x', t') + n_2(x', t') + \dots, \quad (4.151)$$

in which  $n_0$  corresponds to the QS density, while  $n_1, n_2$  etc. represent first-order, second-order etc. NQS components. Obviously, at “sufficiently” slow temporal changes the NQS components are relatively small corrections to the QS density, and the series converges rapidly. Inserting (4.151) into (4.149) and sorting terms with the same  $\alpha^n$  yields the recursive sequence of diffusion equations

$$0 = D_n \frac{d^2 n_0}{dx'^2} \quad (4.152a)$$

$$\frac{\partial n_0}{\partial t} = D_n \frac{d^2 n_1}{dx'^2} \quad (4.152b)$$

$$\frac{\partial n_1}{\partial t} = D_n \frac{d^2 n_2}{dx'^2} \dots \quad (4.152c)$$

As shown in [52], after setting  $\alpha = 1$  the solution can be written as

$$n(x', t') = \Xi_0(x')n_e(t) + \Xi_1(x')\frac{\partial n_e}{\partial t} + \Xi_2(x')\frac{\partial^2 n_e}{\partial t^2} + \dots; \quad (4.153)$$

i.e. the components can be separated into a purely spatially dependent function  $\Xi$  and a purely time-dependent expression. The latter appears in the  $k^{\text{th}}$ -order NQS component as  $k^{\text{th}}$ -order time derivative of the carrier density at the beginning of the neutral base. According to (4.150) these derivatives are given by the time dependence of the controlling (input) voltage. For a short base the spatial functions of the first two components read [52]

$$\Xi_0(x') = 1 - \frac{x'}{w_B}, \quad (4.154)$$

$$\Xi_1(x') = \tau_{IT} \left[ -2 \frac{x'}{w_B} + 3 \left( \frac{x'}{w_B} \right)^2 - \left( \frac{x'}{w_B} \right)^3 \right], \quad (4.155)$$

with the time constant

$$\tau_{IT} = \frac{1}{3} \frac{w_B^2}{2D_n} = \frac{\tau_{Bfd}}{3}. \quad (4.156)$$

The terminal currents are given by the transport equation, taking into account the various simplifications made at the beginning:

$$i_{nC} = -j_n(w, t)A_E = -A_E \frac{qD_n}{w_B} \left[ -n_e(t) + \tau_{IT} \frac{\partial n_e}{\partial t} \right], \quad (4.157a)$$

$$i_{nE} = -j_n(0, t)A_E = -A_E \frac{qD_n}{w_B} \left[ -n_e(t) - 2\tau_{IT} \frac{\partial n_e}{\partial t} \right]. \quad (4.157b)$$

As a result, these two currents consist of the QS transfer current  $i_T(t) = A_E q D_n n_e(t) / w_B$  and a NQS correction term. For a small sinusoidal signal, the collector current can be written in frequency domain as

$$I_{nC} = I_{T0}[1 - j\omega\tau_{IT}] \approx I_{T0}\exp(-j\omega\tau_{IT}). \quad (4.158a)$$

The last expression results from an expansion that is often found in the literature as mentioned earlier. While the  $\exp()$  term clearly corresponds to a signal delay with  $\tau_{IT}$ , it is only valid for sufficiently low frequencies and, hence, should not be used in a compact model.

The dynamic base current is given by

$$i_{pB} = i_{nE} - i_{nC} = A_E \frac{qD_n}{w_B} 3\tau_{IT} \frac{\partial n_e}{\partial t}.$$

It does not contain a DC component since recombination was neglected. With (4.139) for the time dependent QS base charge and (4.156), the base current can be written in the better known form

$$i_B = \frac{\partial Q_{Bf0}}{\partial t}, \quad (4.159)$$

with  $Q_{Bf0} = Q_{nB}(\zeta \rightarrow 0)$  at low injection. The result corresponds to a purely reactive current without additional delay. However, by taking the next NQS term  $n_2(x', t)$  an additional time delay  $\tau_{Qf} = \tau_{Bfd}/6$  is obtained for the base current. Then, also the collector and emitter current become second-order differential equations in time.

The solution procedure described above is valid for arbitrary input signals  $v_{B'E'}$  in time domain as long as the initial assumptions for low injection are met. Along these lines, a detailed analysis was performed in [52] using a controlled current source for each NQS component, which was implemented in parallel to the respective DC element for base and collector current. The impact of the first 3 NQS components on the delay of the time dependent base and collector current were nicely demonstrated. However, despite a smooth transient signal for  $v_{B'E'}$  (rather than an often found unrealistic step-function), it was observed that even “*the four terms used to calculate the currents are not enough to obtain accurate results at all values of time*”, especially at the beginning of the transient. Also, the evaluation of the additional NQS current sources in the equivalent circuit requires the calculation of higher order derivatives of the controlling voltages (e.g.  $v_{B'E'}$ ). Overall, while the approach is very useful for obtaining theoretical insights, it is neither reliable enough nor does it cover the necessary bias

range for production compact models. The high-injection solution presented in [53, 54] does not change this, since conductivity modulation in the base region is usually outside the practical operating range.

Since experimentally the impact of NQS effects can only be detected in the terminal currents, an accurate description of the time-dependent terminal currents is highly desirable. Since the dynamic portion of the currents is defined by the time derivative of the charges, such a solution would in fact also define the location and values of the charge storage elements in an equivalent circuit. The most general way for obtaining time-dependent terminal currents is to take moments of the continuity and transport equation [55, 56]. The first moment leads to the Transient Integral Charge-Control Relation (TICCR), which is selected here as illustrative example. It is practically oriented and sufficient to demonstrate the basic idea as opposed to pursuing a general mathematical approach.

Reorganizing the electron transport equation the same way as for the GICCR (cf. eq. (4.36)) and then performing an integration by parts yields the following equations for the electron current densities at the internal collector and emitter contact [57]:

$$i_C(t) = -j_n(L_x, t)A_E = i_T(t) - qA_E \int_0^{L_x} g_C \left( R + \frac{\partial n}{\partial t} \right) dx, \quad (4.160a)$$

$$i_E(t) = j_n(0, t)A_E = i_T(t) + qA_E \int_0^{L_x} g_E \left( R + \frac{\partial n}{\partial t} \right) dx. \quad (4.160b)$$

$i_T$  is the transfer current expression from the GICCR (4.52) but with  $h_J = 1$ ,

$$i_T(t) = A_E q V_T \mu_{nr} n_{ir}^2 \frac{\exp(V_{B'E'}/V_T) - \exp(V_{B'C'}/V_T)}{\int_0^{L_x} h_t(\xi, t) p(\xi, t) d\xi}, \quad (4.161)$$

and

$$g_C(x, t) = \frac{\int_0^x h_t(\xi, t) p(\xi, t) d\xi}{\int_0^{L_x} h_t(\xi, t) p(\xi, t) d\xi}, \quad (4.162a)$$

$$g_E(x, t) = \frac{\int_x^{L_x} h_t(\xi, t) p(\xi, t) d\xi}{\int_0^{L_x} h_t(\xi, t) p(\xi, t) d\xi} \quad (4.162b)$$

are composite weight functions in which

$$h_t(\xi, t) = h_g(\xi, t) h_e(\xi, t) \quad (4.163)$$

consists of the same weight functions  $h_g$  and  $h_e$  as defined already in the GICCR (cf. (4.37)). Figure 4.43 visualizes the spatial dependence of the TICCR weight function  $g_C$ . The function changes rapidly in the base region, while it is mostly constant in the emitter and collector region. Also, it remains bias independent at lower injection and only starts to change on the emitter side towards higher current densities. Notice, that from (4.162a,b) one obtains

$$g_E = 1 - g_C. \quad (4.164)$$

This relation will be used later again and can also be shown to be valid generally for other moments [55].

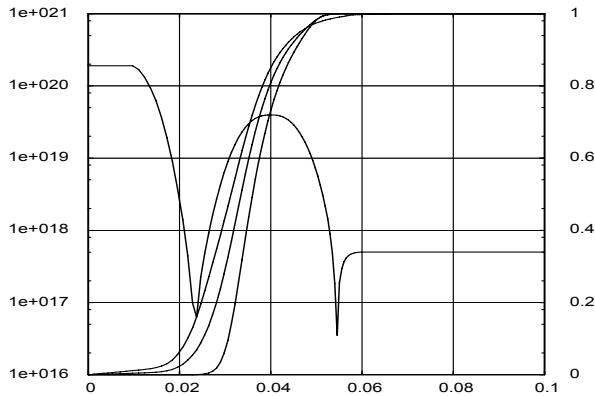


Fig. 4.43: Spatial dependence of weight function  $g_C$  in the CED HBT for selected bias points  $J_C/(\text{mA}/\mu\text{m}^2) = 2.14 \cdot 10^{-3}, 1.94, 5.73$  @  $V_{B'E}/V = 0.7, 0.885, 0.92$  and  $V_{B'C} = 0V$ . The dotted line shows the doping profile for reference.

For practical applications and the intended investigation of high-speed transients the recombination rate  $R$  can be neglected in (4.160a,b). This is

also consistent with setting the GICCR function  $h_J = 1$  in  $i_T$ . The resulting expressions are compared with actual device simulation data in Fig. 4.44 for nonlinear large-signal transient operation. The transistor is biased at peak  $f_T$ . A signal voltage  $v_{B'E'}$  with 50mV amplitude drives the transistor significantly into the high-current region with large charge storage and a bias dependent transit and delay time. The resulting nonlinear response of  $i_C(t)$  is excellently described by (4.160a) which, therefore, can be used as reference for further model development. Figure 4.44 also shows the delay of  $i_C(t)$  w.r.t. both  $v_{B'E'}$  and  $i_T(t)$ .

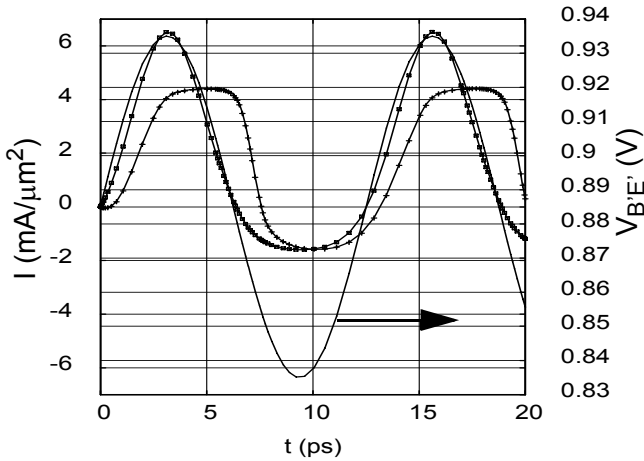


Fig. 4.44: Time-dependent terminal current of the CED HBT with 50 mV amplitude biased at  $J_C = 1.94 \text{ mA}/\mu\text{m}^2$ ,  $V_{B'E'} = 0.885 \text{ V}$ ,  $V_{B'C'} = 0 \text{ V}$ . Comparison of TICCR (solid lines) to the exact solution from device simulation (crosses). The recombination contribution was neglected in the analytical equation. For reference, the QS current  $i_T(t)$  is inserted (dashed line).

To make the TICCR suitable for compact modeling the time derivative term needs further inspection. As it turns out, the change of  $g_C$  and  $g_E$  with time is much smaller than that of the electron density. Applied to the discrete representation in transient simulation, this means that the time derivative of  $g(x,t)$  is negligible from *one time step to the next*. This allows the time derivative operator in (4.160a,b) to be moved in front of the integral leading to



$$qA_E \int_0^{L_x} g(x, t) \frac{\partial n(x, t)}{\partial t} dx \cong \frac{\partial}{\partial t} \left( qA_E \int_0^{L_x} g(x) n(x, t) dx \right) \quad (4.165)$$

within a given time interval.

The expression in the parenthesis on the r.h.s. of (4.165) represents a weighted charge. Hence, the two weight functions define a charge associated with the collector terminal,

$$Q_{tC}(t) = qA_E \int_0^{L_x} g_C(x) n(x, t) dx, \quad (4.166a)$$

and a charge associated with the emitter terminal,

$$Q_{tE}(t) = qA_E \int_0^{L_x} g_E(x) n(x, t) dx. \quad (4.166b)$$

With these definitions and  $R = 0$  eqs. (4.160a,b) reduce to

$$i_C(t) = i_T - \frac{\partial Q_{tC}(t)}{\partial t} \quad (4.167a)$$

$$i_E(t) = i_T + \frac{\partial Q_{tE}(t)}{\partial t}. \quad (4.167b)$$

Physically, during dynamic operation above equations define the portion of stored (weighted) charge in the transistor that flows through the collector terminal and the emitter terminal, respectively. In other words, the detailed evaluation of the two charge components  $Q_{tC}$  and  $Q_{tE}$  may be used to define the *location* of the charge storage elements in the EC. This is in contrast to quasi-static transistor theory in which the charge storage elements are allocated to the nodes empirically.

At this point it is also worthwhile to calculate the base current,

$$i_B(t) = i_E(t) - i_C(t) = \frac{\partial Q_{tE}(t)}{\partial t} + \frac{\partial Q_{tC}(t)}{\partial t} = \frac{\partial Q_p(t)}{\partial t}. \quad (4.168)$$

where (4.165) was used to obtain the most right expression. As expected,  $i_B$  supports the total change of charge  $\partial Q_p = \partial Q_n$  in the transistor. Generally, any NQS behavior in the terminal currents is caused by an NQS behavior of the charges.

With the earlier definition (GICCR) for the excess minority carrier density the weighted charges, related to either emitter or collector, can be split into a depletion component (index  $j$ ) and a minority component (index  $m$ ):

$$\Delta Q_t(t) = \Delta Q_{t,j}(t) + \Delta Q_{t,m}(t). \quad (4.169)$$

Hence, the TICCR depletion charges are given by

$$\Delta Q_{tC,j}(t) = A_E q \int_{x(|\delta p| < |\delta n|)} g_C(\Delta n - \Delta p) dx, \quad (4.170a)$$

$$\Delta Q_{tE,j}(t) = A_E q \int_{x(|\delta p| < |\delta n|)} g_E(\Delta n - \Delta p) dx. \quad (4.170b)$$

Taking a closer look at the weight functions in Fig. 4.43 reveals that  $g_E$  and  $g_C$  are equal to 1 within the regions of interest; i.e. within the emitter sided BE SCR and the collector sided BC SCR. Thus,

$$\Delta Q_{tC,j}(t) = Q_{jC}(t), \quad (4.171a)$$

$$\Delta Q_{tE,j}(t) = Q_{jE}(t). \quad (4.171b)$$

For the same reason, the weighted minority charge of the collector and emitter region are simply given by

$$\Delta Q_{tC,m}(t) = Q_{pC}(t), \quad (4.172a)$$

$$\Delta Q_{tE,m}(t) \cong Q_{pE}(t). \quad (4.172b)$$

The “approximate” sign holds at higher current densities, when  $g_C$  ( $g_E$ ) deviates from 0 (1) in the metallurgical emitter.

However, the situation is very different for the remaining minority charge component in the base. Due to the variation of the weight functions across the base region only a certain portion of the stored minority charge is included in  $\Delta Q_{tE,m}(t)$  while the remaining fraction is contained in  $\Delta Q_{tC,m}(t)$ . With (4.164) one can write for the weighted TICCR minority charges

$$\Delta Q_{tC,mB}(t) = \alpha_t Q_{mB}(t), \quad (4.173a)$$

$$\Delta Q_{tE,mB}(t) = (1 - \alpha_t) Q_{mB}(t) \quad (4.173b)$$

Here,  $\alpha_t$  is the base charge partitioning factor,

$$\alpha_t = \frac{\Delta Q_{tC, mB}(t)}{\Delta Q_{mB}(t)} = \frac{qA_E \int_{x_{iE}}^{x_{jC}} \Delta(g_C(x)n(x, t))dx}{qA_E \int_{x_{iE}}^{x_{jC}} \Delta n(x, t)dx}, \quad (4.174)$$

which depends on bias and time; i.e. even for identical bias points during transient analysis, the factor can be different due to the NQS behavior of the carrier density. An evaluation of the above expressions requires the knowledge of the carrier densities.

Inserting above charges into (4.167a,b) leads to the currents

$$i_C(t) = i_T - \frac{\partial Q_{jC}}{\partial t} - \frac{\partial Q_{pC}}{\partial t} - \frac{\partial \Delta Q_{tC, mB}}{\partial t} \quad (4.175a)$$

$$i_E(t) = i_T + \frac{\partial Q_{jE}}{\partial t} + \frac{\partial Q_{pE}}{\partial t} + \frac{\partial \Delta Q_{tE, mB}}{\partial t}. \quad (4.175b)$$

The delay is mainly caused by the last term, although both  $Q_{pC}$  (at high current densities) and  $Q_{pE}$  can contain NQS components via integration of the corresponding NQS carrier density portion. For advanced transistors, however, the emitter is very shallow allowing to neglect NQS effects in  $Q_{pE}$ . Furthermore, in SiGe HBTs the injection of minority carriers into the emitter and collector only occurs at *very* high current densities, which are of little practical importance. Hence, the impact of any NQS effects in  $Q_{pE}$  and  $Q_{pC}$  can usually be neglected.

In order to obtain approximate analytical expressions for  $\Delta Q_{t, mB}$  and the partitioning factor, a 1D diffusion transistor operating at sufficiently low injection is assumed. Thus, the weight functions read

$$g_C(x) = x/w_B, \quad g_E(x) = 1 - x/w_B. \quad (4.176)$$

Inserting the first two terms of the minority carrier density profile (4.153) with  $\mathcal{E}$  from (4.154) and (4.155) gives

$$\Delta Q_{tC, mB}(t) = \frac{Q_{nB0}(t)}{3} - \frac{7}{30} \tau_{IT} \frac{\partial Q_{nB0}(t)}{\partial t}, \quad (4.177)$$

$$\Delta Q_{tE, mB}(t) = 2 \frac{Q_{nB0}(t)}{3} + \frac{22}{30} \tau_{IT} \frac{\partial Q_{nB0}(t)}{\partial t}. \quad (4.178)$$

Each of the weighted transient charges contains a QS portion and a (first-order) NQS component which is given by the time derivative of the QS portion and a time constant. The charge partitioning factor (4.174) is about 1/3 in this case. Note that according to low-injection theory  $w_B$  is not assumed to be time dependent. Also,  $Q_{pC}$  is negligible so that

$$i_C(t) = i_T - \frac{\partial Q_{jC}}{\partial t} - \frac{1}{3} \frac{\partial Q_{nB0}}{\partial t} + \frac{7}{30} \tau_{IT} \frac{\partial^2 Q_{nB0}}{\partial t^2}, \quad (4.179a)$$

$$i_E(t) = i_T + \frac{\partial Q_{jE}}{\partial t} + \frac{\partial Q_{pE}}{\partial t} + \frac{2}{3} \frac{\partial Q_{nB0}}{\partial t} + \frac{22}{30} \tau_{IT} \frac{\partial^2 Q_{nB0}}{\partial t^2}. \quad (4.179b)$$

This means that roughly 1/3 of the base minority charge is “reclaimable” by the collector and the remaining 2/3 by the emitter. The second-order derivative terms lead mostly to a magnitude change and are important for the NQS behavior of the dynamic base current, which reads at low injection with the above results

$$i_B(t) = \frac{\partial Q_{jE}}{\partial t} + \frac{\partial Q_{jC}}{\partial t} + \frac{\partial Q_{pE}}{\partial t} + \frac{\partial Q_{nB0}}{\partial t} + \tau_{QmB} \frac{\partial^2 Q_{nB0}}{\partial t^2}. \quad (4.180)$$

Only the last term represents the NQS effect with the characteristic time

$$\tau_{QmB} = \frac{1}{2} \tau_{IT} = \frac{1}{6} \tau_{fBd}. \quad (4.181)$$

which is equivalent to the “delay” time.

In the literature and most models only the QS carrier distribution  $\mathcal{E}_0$  from (4.154) is inserted, which produces only a first-order description of the terminal currents. This approach corresponds to the well-known partitioned-charge based (PCB) model [62, 54] in which only the additional phase of the transfer current is considered while the minority charge description remains quasi-static. The consequences of this simplification will be demonstrated later.

The results of the two time-domain approaches presented above can be summarized as follows. Taking moments of the transport or continuity equation yields a general solution for the state variables, i.e. the currents at the collector and emitter contact, as a function of time for arbitrary shapes of signals. However, in order to be able to obtain analytical expressions

suitable for compact models the carrier density (and some other internal variables) need to be known. For this, the time series approach can be used to calculate the carrier density profile for arbitrary large-signal transients. Using the QS and first-order NQS term of the time series solution leads to NQS current components for the collector and base current. The characteristic time constants associated with these NQS components can be identified as delay times.

#### 4.5.3 Frequency domain solution

If the considerations are restricted to small signals, (4.148) can be linearized in a given operating point and then be transformed into frequency domain, yielding

$$j\omega \underline{n} = \bar{\mu}_{nB} V_T \frac{d^2 \underline{n}}{dx^2} + \bar{\mu}_{nB} \bar{E}_n \frac{d \underline{n}}{dx} \quad (4.182)$$

with the small-signal electron density  $\underline{n}$  as complex variable. Although this is also a second-order differential equation, it is now much easier to solve since mobility and field are constants due to the bias point linearization. Note that (4.182) only holds for sufficiently low injection levels.

For neutral regions a closed-form solution of (4.182) can be obtained, which includes the influence of the drift field and was presented first in [50, 58]. In the following discussion the neutral base region is considered with the emitter side defined at  $x = 0$  and the collector side at  $x = w_B$ . Then the corresponding solution for the frequency and spatially dependent electron density reads in normalized form

$$\frac{\underline{n}(x, \omega)}{\underline{n}_{e0}} = \exp\left(\frac{\zeta x}{2w_B}\right) \frac{\sinh([1 - x/w_B]\Lambda)}{\sinh(\Lambda)}. \quad (4.183)$$

Here,  $\zeta$  is the drift factor and

$$\Lambda = \sqrt{\left(\frac{\zeta}{2}\right)^2 + j\Omega} = \frac{\zeta}{2} \sqrt{1 + j4\Omega\zeta^{-2}}, \quad (4.184)$$

is a complex variable of the normalized frequency

$$\Omega = \omega \frac{w_B^2}{\bar{\mu}_{nB} V_T} = \omega \tau_{Bfd} F_\zeta, \quad (4.185)$$

which is expressed by the drift-diffusion portion (3.174) of the base transit time with  $F_\zeta$  from (3.175). Finally, the quasi-static small-signal carrier density at the emitter side of the base is obtained from (3.48) with  $n_c = 0$ ,

$$n_{e0} = J_{T0} \frac{w_B}{q \bar{\mu}_{nB} V_T} \frac{1 - \exp(-\zeta)}{\zeta}, \quad (4.186)$$

with the quasi-static small-signal transfer current density

$$J_{T0} = \frac{I_T}{A_E V_T} V_{B'E} \quad (4.187)$$

and  $I_T$  as DC transfer current.

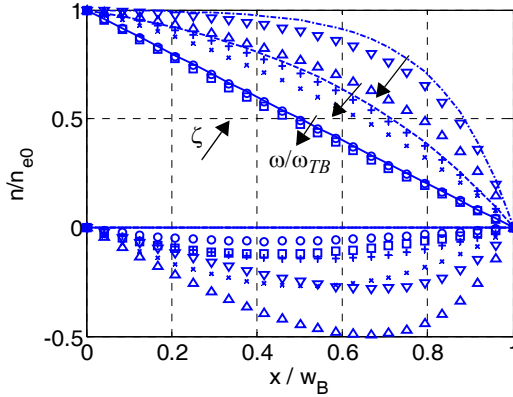


Fig. 4.45: Visualization of the real and imaginary part of the electron density from (4.183) along the normalized base width for  $\zeta = \{0, 2, 6\}$  and  $\omega/\omega_{TB} = \{0, 0.5, 1\}$ . The lines and symbols of the real and imaginary part correspond to each other.

The obtained solution (4.183) is visualized in Fig. 4.45 for different values of the drift factor and normalized frequency. The normalization factor  $\omega_{TB} = 1/(2\pi\tau_{Bfd})$  corresponds to the transit frequency of the neutral base region (with  $n(w_B) = 0$ ). For the real part (and magnitude) deviations from the QS distribution can be observed beyond  $\omega/\omega_{TB} = 0.5$  and increase with  $\zeta$ . However, the imaginary component is visibly non-zero already below  $\omega/\omega_{TB} = 0.5$ , indicating a pronounced phase shift. The higher the drift

field the larger is the imaginary part (and thus the phase shift) with its maximum shifted towards the end of the base.

Assuming  $n(w_B) = 0$ , the transfer current follows from the diffusion current at  $w_B$ :

$$I_T = -A_E q \bar{\mu}_{nB} V_T \frac{d\bar{n}}{dx} \Big|_{w_B} = \frac{A_E q \bar{\mu}_{nB} V_T}{w_B} n_{e0} \frac{\Lambda}{\sinh(\Lambda)} \exp\left(\frac{\zeta}{2}\right). \quad (4.188)$$

Inserting (4.186) yields for the normalized frequency dependent transfer current

$$\frac{I_T(\omega)}{I_{T0}} = \frac{\sinh(\zeta/2)}{\zeta/2} \frac{\Lambda}{\sinh(\Lambda)}. \quad (4.189)$$

The frequency dependence in Fig. 4.46 shows that the reduction in magnitude and the increase in phase increase with drift factor.

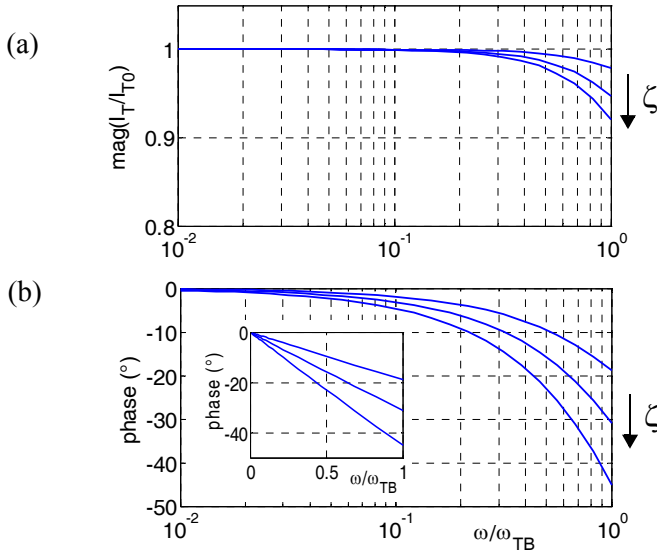


Fig. 4.46: Small-signal transfer current vs. normalized frequency: (a) normalized magnitude and (b) phase. Evaluation of exact solution (4.191) with drift factor  $\zeta = 0, 2, 6$  as parameter. The inset shows the phase vs. a linear frequency axis.

From (4.183), the small-signal current *injected into the base* at  $x = 0$  is

$$I_E(\omega) = \frac{A_E q \bar{n}_{nB} V_T}{w_B} n_{e0} \frac{\Lambda \cosh(\Lambda) - \frac{\zeta}{2} \sinh(\Lambda)}{\sinh(\Lambda)}, \quad (4.190)$$

from which the small-signal transport factor can be calculated,

$$\alpha_T = \frac{I_T(\omega)}{I_E(\omega)} = \frac{\Lambda \exp\left(\frac{\zeta}{2}\right)}{\Lambda \cosh(\Lambda) - \frac{\zeta}{2} \sinh(\Lambda)}, \quad (4.191)$$

which will be needed during the discussion on noise. Its zero-field expression simply reads

$$\alpha_T|_{\zeta=0} = \frac{1}{\cosh(\Lambda)} \quad (4.192)$$

and is well-known from classical transistor theory. It corresponds to the small-signal current gain (of the neutral base region) in common-base configuration.

The small-signal charge is obtained from integrating the carrier density (4.183) over the neutral base region:

$$\underline{Q}_{nB} = \frac{q A_E n_{e0} w_B}{j\Omega} \frac{\frac{\zeta}{2} \sinh(\Lambda) + \Lambda \cosh(\Lambda) - \Lambda e^{\zeta/2}}{\sinh(\Lambda)}. \quad (4.193)$$

With the quasi-static value

$$\underline{Q}_{nB0} = q A_E n_{e0} w_B \frac{\frac{\zeta}{2} e^{\zeta/2} - \sinh(\zeta/2)}{\zeta \sinh(\zeta/2)} = \tau_{Bfd} I_{T0} \quad (4.194)$$

one obtains for the normalized frequency dependent charge

$$\frac{\underline{Q}_{nB}}{\underline{Q}_{nB0}} = \frac{\zeta \sinh\left(\frac{\zeta}{2}\right)}{j\Omega} \frac{\frac{\zeta}{2} \sinh(\Lambda) + \Lambda \cosh(\Lambda) - \Lambda e^{\zeta/2}}{\frac{\zeta}{2} e^{\zeta/2} - \sinh\left(\frac{\zeta}{2}\right) \sinh(\Lambda)}. \quad (4.195)$$



The frequency dependence in Fig. 4.47 shows that the magnitude decreases slightly with the drift factor. In contrast, the phase shift increases rapidly but is only about half of that of the transfer current.

Under the assumed simplifying conditions above formulations for current and charge are exact solutions. However, they are not suitable yet for use in compact models since there is no equivalent representation in time domain. Considering the fact that the high frequency characteristics of a transistor are strongly affected by external parasitics it is permissible for practical purposes to approximate the exact solutions for the intrinsic transistor by their lower order terms. Therefore, a series expansion is performed from which approximate solutions are derived by truncation after the second-order frequency term.

Series expansion gives for the transfer current

$$\frac{I_T(\omega)}{I_{T0}} \approx [1 - (\omega\tau_{IT,2})^2] - j\omega\tau_{IT,1} \quad (4.196)$$

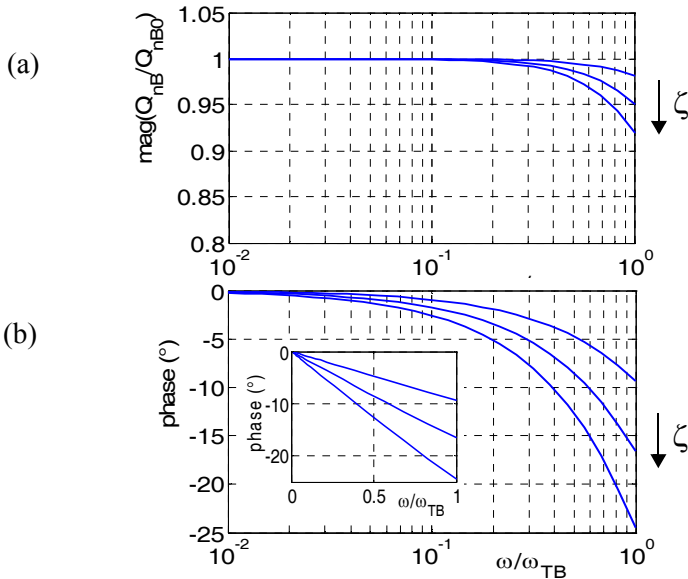


Fig. 4.47: Small-signal (base) minority charge vs. normalized frequency: (a) normalized magnitude and (b) phase. Evaluation of exact solution (4.196) with drift factor  $\zeta = 0, 2, 6$  as parameter. The inset shows the phase vs. a linear frequency axis.

The time constants can also be expressed as

$$\tau_{IT,1} = g_{IT,1} \tau_{Bfd} \quad \text{and} \quad \tau_{IT,2} = g_{IT,2} \tau_{Bfd} \quad (4.197)$$

in which the coefficients  $g_{IT}$  are complicated functions of the drift field (e.g. [58]). According to this result the phase shift can be related to a characteristic time constant that is proportional to  $\tau_{Bfd}$ . For the special case of a diffusion transistor ( $\zeta = 0$ ), one obtains

$$\tau_{IT,1} = \tau_{Bfd}/3 \quad (4.198)$$

The first-order time constant corresponds to three times the neutral base transit frequency  $f_{TB}$  and causes a phase shift at  $f_{TB}$  of 18 degrees as can be seen in Fig. 4.46. Physically, the above result can be interpreted such that the fraction  $g_{IT,1}$  (i.e. 1/3 in case of a diffusion transistor) of the stored base charge is retrieved through the collector terminal. By calculating the corresponding emitter current it can be shown that the remaining fraction ( $1 - g_{IT,1}$ ), which is 2/3 for a diffusion transistor, is retrieved through the emitter current.

For the minority charge, series expansion leads to the approximation

$$\frac{\underline{Q}_{nB}}{\underline{Q}_{nB0}} = [1 - (\omega \tau_{QB,2})^2] - j\omega \tau_{QB,1}, \quad (4.199)$$

in which the characteristic time constants can be expressed as a function of the base transit time

$$\tau_{QB,1} = g_{QB,1} \tau_{Bfd} \quad \text{and} \quad \tau_{QB,2} = g_{QB,2} \tau_{Bfd}. \quad (4.200)$$

Again, the coefficients  $g_{QB}$  are fairly complicated functions of the drift field. For the special case of a diffusion transistor ( $\zeta = 0$ ), one obtains for first-order coefficient that determines the phase shift:

$$\tau_{QB,1} = \tau_{Bfd}/6. \quad (4.201)$$

The NQS factors  $g_{IT}$  and  $g_{QB}$  increase with drift field. For instance, in the limiting case of an infinite drift field,  $g_{IT,1} = 1$  and  $g_{QB,1} = 0.5$ . In practice, the NQS factors  $g_{IT}$  and  $g_{QB}$  can be determined from the *excess* phase shift and magnitude drop of  $y_{21}$  and  $y_{11}$ . “Excess” here means that the phase shift and magnitude changes resulting from other effects need to be already taken into account. For instance,  $\omega C_{jC}$  and the input time con-

stant  $r_{Bi}C_{BEi}$ , respectively, also cause a phase shift and a magnitude drop, respectively, in  $y_{2I}$ .

There have been various attempts at extending classical theory. The most noticeable one [59] was the derivation of a more general solution that includes the impact of the spatially dependent mobility and intrinsic carrier density due to high-doping effects. Also, the solution approach is based on a power series in frequency for the electron density, which is similar to the approach outlined earlier in this chapter for the time domain. As a result, the QS as well as the higher order NQS components of the carrier density and derived variables are obtained. In addition, the charge partitioning factor is expressed more generally, and the solution is also applied to the neutral emitter taking into account the effect of a finite emitter contact recombination on the dynamic base current. While this and other extensions provide insight into the principle of internal transistor operation, their value for compact modeling is limited since the regional parameters and detailed information on the doping profile are often not available in practice. Also, in most HBTs the impact of the neutral emitter charge becomes very small due to the injection barrier and the extremely shallow junction.

More general results can also be obtained from the TICCR. Since the time dependence of  $g_E$  and  $g_C$  can be neglected for small-signal operation, eqs. (4.166a,b) can be easily transformed into frequency domain, yielding

$$\underline{Q}_{tC}(\omega) = qA_{E0} \int_0^{L_x} g_C(x) \underline{n}(x, \omega) dx \quad (4.202a)$$

$$\underline{Q}_{tE}(\omega) = qA_{E0} \int_0^{L_x} g_E(x) \underline{n}(x, \omega) dx, \quad (4.202b)$$

where  $\underline{n}(x, \omega)$  is the small-signal frequency dependent carrier density, and  $g_E$  and  $g_C$  are calculated from the d.c. solution. This allows a fast evaluation of the partitioning factor  $\alpha_t$  defined in (4.174).

The results of the frequency-domain approach above can be summarized as follows. After linearization and under simplified assumptions, the combined continuity and transport differential equation can be solved directly in closed-form for its state variable, i.e. the electron density in the neutral base region. Postprocessing then yields the transfer current and charge as a function of frequency. Proper truncation of the resulting infinite series

lead to a QS and a NQS component for the collector and base current. The phase shift terms in those expressions contain characteristic time constants that are identical with those of the time-domain theory. In fact, frequency-domain theory leads for those cases, in which closed-form analytical solutions can be obtained, to the same results as the TICCR and its extensions. It is not necessary to resort to mysterious “delay times” as in [58] in order to justify the theory and its results. Also, the resulting and often used  $\exp(-j\omega\tau)$  expression does not have a theoretical basis. Nevertheless, the frequency-domain results are not suitable for a compact model, unless a consistent representation in time domain can be found.

#### 4.5.4 Discussion and relation to compact modeling

The evaluation so far has been performed for low injection. However, at practically relevant medium current densities, when the region boundaries, mobility, and electric field become a nonlinear function of bias and carrier density no closed-form solutions are known. It seems that for the general case of large-signal terminal voltages with arbitrary shape and transistor operation at arbitrary injection levels a closed-form solution is not possible. Therefore, the theory derived above can only be used as a guideline for constructing a compact model. This results in different approaches and implementations, which need to be verified against device simulation and measurements.

The main task for a compact model is to arrive as a *consistent* formulation for time- *and* frequency-domain simulation. As shown in the literature, in both cases the solution form at very high injection is the same as for low-injection, just the parameters are different. Hence, it is reasonable to assume that it can also be used in between these limiting cases. Both time and frequency solution approaches also have the first- and second-order time derivative and  $j\omega$  term, respectively, in common. Converting the frequency-domain solution  $\underline{X}$  ( $= I_T$  or  $Q_j$ ) to a slightly different form (with  $s = j\omega$ ),

$$\frac{\underline{X}_{nqs}(\omega)}{\underline{X}} = 1 - s\tau_1 + (s\tau_2)^2 \cong \frac{1}{1 + A_1s + A_2s^2}, \quad (4.203)$$

which is valid for not too high frequencies, allows to apply Laplace transformation in order to obtain the corresponding time-domain approximation,

$$A_2 \frac{d^2 x_{nqs}}{dt^2} + A_1 \frac{dx_{nqs}}{dt} + x_{nqs} = x(t), \quad x = (i_T, Q_f). \quad (4.204)$$

This transformation is always valid as long as the linearized (small-signal) case is also considered in time-domain. The form of the obtained second-order differential equation resembles the existing time-domain solution. As shown in [60] the above system accomplishes in both domains the task of properly shifting the terminal currents as a consequence of NQS effects if the proper delay times are inserted.

One might argue whether the quadratic frequency term in the approximate current and charge expressions is really necessary. Thus, it is instructional to compare the approximations with the results obtained from device simulation with a realistic doping profile. Figure 4.48 contains the corresponding frequency dependent small-signal collector current (i.e. transconductance) and base current (i.e. input admittance). Note, that the impact of  $C_{jCi}$  on the results is negligible at the selected bias point. The conversion (4.203) along with the second-order term causes the magnitudes of both input admittance and transconductance to follow the physically correct drop. The form  $(1 + A_1 s + A_2 s^2)^{-1}$  is in fact more accurate than the original form (middle term) in (4.203) over a wider frequency range, although it was realized here as Bessel polynomial in order to be able to approximate the second-order time constant by the first-order one. In other words, the conversion apparently leads to a partial error compensates and allows to ignore higher order terms of the original solution. This example shows that the above transformation system not only is consistent in frequency- and time-domain but also yields accurate results. Also, as will be seen later, the polynomial form of the approximate solution proves to be useful for a representation in compact models.

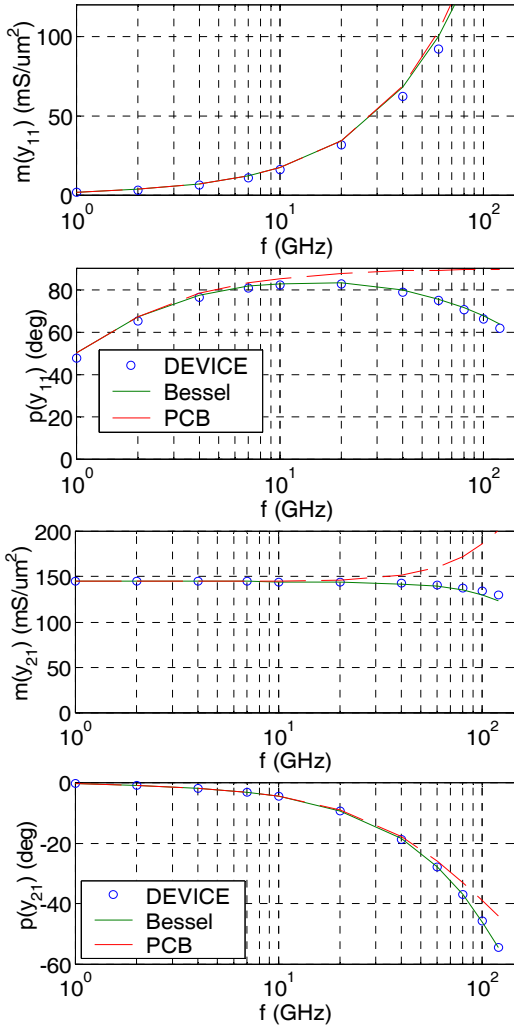


Fig. 4.48: Input admittance (upper figures) and transconductance (lower figures) of the CED HBT operated at  $J_C = 6.1 \text{ mA}/\mu\text{m}^2$ ,  $V_{BE'} = 0.922 \text{ V}$ ,  $V_{BC'} = 0 \text{ V}$ . Comparison of  $(1 + A_1 s + A_1 s^2)^{-1}$  from (4.203) (solid lines) and the PCB approach (dashed lines) with AC device simulation results (symbols).

Also shown in Fig. 4.48 is the PCB model. It only allows to model the phase shift of the collector current, while the base current remains QS. In particular the magnitude of the collector current increases non-physically

with frequency. Moreover, as a consequence of ignoring the additional phase of the base current, the (experimentally) observed phase shift in the *current gain* cannot be described correctly. Therefore, this approach and the associated implementation using a transcapacitance are not recommended for compact modeling.

As mentioned before, for general large-signal operation no analytical solution exists. Observation from device simulation and measurements at medium and high-current densities as well as for pulse signals with several 100mV amplitude (e.g. [28, 61] indicates that there is no fundamental difference in the NQS behavior. In other words, the terminal currents react delayed w.r.t. the input voltage at any bias condition.

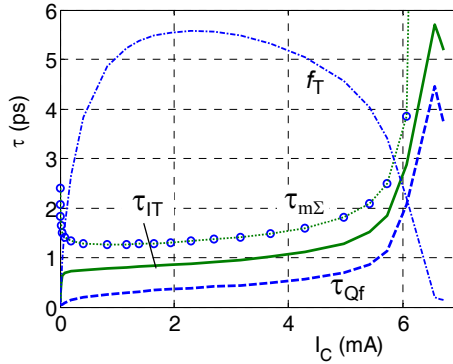


Fig. 4.49: NQS delay times  $\tau_{IT}$  and  $\tau_{Qf}$  vs. bias current for the CED HBT. For reference, the accumulated transit time  $\tau_{m\Sigma}$  from Regional Analysis as well as  $f_T$  have been inserted.

As shown in Fig. 4.49, the delay times, as calculated from device simulation, just increase with transfer current and are fairly well proportional to the total transit time  $\tau_f$ . Thus, the second-order differential equation should still be applicable to large-signal transients for accomplishing the delay. In practice, transient simulation consists of a sequence of circuit equation system solutions at discrete time points. The change between time points is generally small enough to consider the theoretical solution as a sufficiently accurate representation, if just the time constants at each time (i.e. bias) point are inserted. In other words, the linearized delay transformation form (4.204) is applied to each time point. The accuracy of this approach has been demonstrated numerous times by examples ranging from device

simulation to production circuit design. The strong coupling between the delay times and  $\tau_f$  in Fig. 4.49 also indicates that the theoretical approach that was derived so far for the neutral base only should be applicable to the entire transistor even at high current densities.

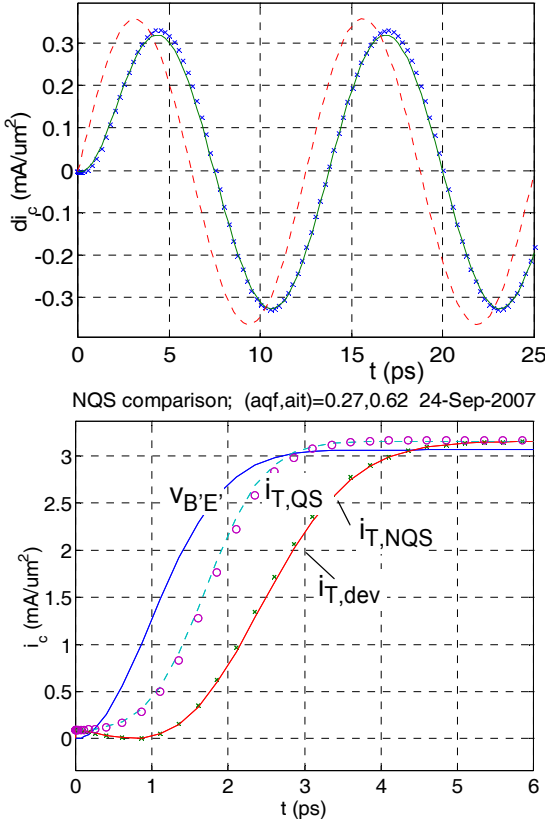


Fig. 4.50: Comparison of (4.204) (solid line) with transient device simulation results (crosses) for the CED HBT. (a) Small-signal sinusoidal input signal with 1mV amplitude; bias point:  $(V_{B'E'}, J_C) = (0.922V, 6.1mA/\mu m^2)$ . (b) Large-signal pulse with 100mV amplitude (normalized  $v_{B'E'}$ ), switching from  $(V_{B'E'}, J_C) = (0.8V, 0.95 \mu A)$  to  $(0.9V, 3mA)$ .  $V_{B'C'} = 0V$ . As reference, the transfer current  $i_{T,QS}$  (dashed line) and in (b) also the TICCQR QS transfer current (4.161) (circles) are inserted.

To validate above assumptions, (4.204) is applied in Fig. 4.50a to a transient simulation with small-signal amplitude, which is just the time domain



representation of the results in Fig. 4.48. For reference, the QS transfer current  $i_{T,QS}$ , calculated from the GICCR with QS carrier densities, is inserted, which follows the signal voltage without delay. Since at the selected bias point the delay times are strongly bias current dependent, any additional derivative term should be visible in the result. However, using the same delay as in from Fig. 4.48 the current calculated from (4.204) is almost on top of the device simulated current. The same result is obtained if a bias dependent delay time at each time point is inserted. This confirms the applicability of (4.204) using delay times taken at the bias point.

As a second step of verification, the comparison is performed for a large-signal pulse with 100 mV amplitude (cf. Fig. 4.50b). Using the QS transfer current  $i_{T,QS}$  for  $x(t)$  in (4.204) and including the term  $dQ_{JC}/dt$  yields excellent agreement with the collector terminal current from device simulation. Note, that at each time (bias) point  $\tau_{IT} \approx \alpha_{IT}\tau_f$  is used with  $\tau_f$  as bias dependent transit time according to Fig. 4.49 but  $\alpha_{IT}$  as constant average value. For comparison, also the QS transfer current  $i_{T,QS}$  calculated from the GICCR (based on the QS carrier distribution) and from the TICCR (time dependent carrier distribution according (4.161)) are inserted. Their difference is quite small amounting to maximal 10% at the steepest slope of the input signal. Hence, for practical applications one can calculate  $i_{T,QS}$  from the GICCR and  $Q_{f,QS}$  from the existing QS equations, and then apply the transformation (4.204) to obtain the respective NQS components.

What is the advantage of the TICCR and its extensions over the classical theory? First of all, the TICCR offers an exact solution of the transport and continuity equation for the *complete* 1D transistor structure if the carrier densities are known. Therefore, they can serve as reference for deriving simplified expressions that are suitable for compact modeling, even under large-signal transient operation. Secondly, their derivation from taking moments of the transport or continuity equation reveals that any weighting function can be used that is suitable for approximating as accurately as possible the actual time or frequency dependence of the terminal currents [55]. However, the TICCR and its extensions cannot be used directly for compact modeling unless an analytical formulation for the weighted charges can be found. Typically, this is only possible under similarly simplifying assumptions as for the classical theory and was shown in the pre-

vious section for the example of a diffusion transistor. In other words, the evaluation of the integrals requires knowledge about the spatial dependence of both the weight functions and the carrier density.

#### 4.6 References

- [1] J. Cressler (ed.), "Silicon heterostructure Handbook", CRC Press NY, 2005.
- [2] J. Slotboom, G. Streutker, A. Pruijmboom, D. Gravesteijn, "Parasitic energy barriers in SiGe HBTs", IEEE Electron Dev. Lett., Vol. 12, pp. 486-488, 1991.
- [3] M. Schröter, H. Tran, W. Kraus, "Germanium profile design options for LEC-SiGe HBTs", Solid-State Electronics, Vol. 48, pp. 1133-1146, 2004.
- [4] J. Matthews and A. Blakeslee, "Defects in epitaxial multilayers", Journal of Crystal Growth, Vol.27, pp. 118-125, 1974, and Vol. 29, pp. 273-280, 1975.
- [5] A. Fischer, "Stability constraints in SiGe epitaxy", Chapter 2.7 (pp. 127-141) in "The SiGe Handbook", ed. by J. Cressler, CRC Press NY, 2005.
- [6] T. Ghani, "Characterization of minority carrier transport and heavy doping effects in  $n_{\text{Si}}/p\text{-SiGe}/n_{\text{Si}}$  heterojunction bipolar transistors", Ph.D. thesis, Stanford University, 1994.
- [7] A. Gruhle, A Schueppen, H. Kibbel, U. Koenig, "Minority carrier mobility and density of states in SiGe HBTs determined by a new method", Proc. 21st Int. Symp. on Compound Semicond., pp. 253-258, 1994.
- [8] T. Manku, A. Nathan, "Valence energy-band structure for strained group-IV semiconductors", J. Appl. Phys. 73, pp. 1205-1213, 1993.
- [9] F.M. Bufler, P. Graf and B. Meinerzhagen, "Hole transport investigation in unstrained and strained SiGe", J. Vac. Sci. Tech., Vol. 16, pp. 1667-1669, 1998.
- [10] S. Tiwari, "A new effect at high currents in heterostructure bipolar transistors", IEEE Trans. Electron Dev., Vol. 9, pp. 142-144, 1988.
- [11] B. Mazhari and H. Morkoc, "Effect of collector-base valence band discontinuity on Kirk effect in double heterojunction bipolar transistors", Applied Physics Letters, Vol. 59, No 17., pp/ 2162-2164, Oct. 1991.

- [12] J. Song, J. Yuan, "Modeling the base collector heterojunction barrier effect at high current densities of SiGe HBTs", *Solid-State Electron.*, Vol. 43, pp. 457-461, 1999.
- [13] K. Roenker, P. Mushini, "Dynamic formation of a parasitic barrier to electron flow in SiGe HBTs operating at high current densities", *Microelectronics J.*, Vol. 31, pp. 353-358, 2000.
- [14] R. Kumar and L.P. Hunter, "Collector capacitance and high-level injection effects in bipolar transistors", *IEEE Trans. Electron Dev.*, Vol. ED-22, pp. 51-60, 1975.
- [15] S. Wilms and H.-M. Rein, "Analytical high-current model for the transit time of SiGe HBTs", *Proc. BCTM*, pp. , 1998.
- [16] M. Schröter, H. Tran, "Modeling of charge and collector field in Si-based bipolar transistors", *Proc. WCM, International NanoTech Meeting, Boston (MA)*, pp. 102-107, 2004.
- [17] M. Schröter, M. Friedrich, and H.-M. Rein, "A generalized Integral Charge-Control Relation and its application to compact models for silicon based HBT's", *IEEE Trans. Electron Dev.*, Vol. 40, pp. 2036-2046, 1993.
- [18] M. Schröter, H. Tran, "Two-/three-dimensional GICCR for Si/SiGe bipolar transistors" *Proc. WCM, International NanoTech Meeting, Anaheim (CA)*, pp. 99-104, 2005.
- [19] M. Schröter, "Integral Charge-Control Relations" Chapter A3 (pp. 1181-1208) in "The SiGe Handbook", ed. by J. Cressler, CRC Press NY, 2005.
- [20] H.N. Ghosh, F.H. De La Moneda, N.R. Dono, "Computer-aided transistor design, characterization, and optimization", *Proc. IEEE*, Vol. 55, pp. 1897-1912, 1967.
- [21] H. Krömer, "Two-integral relations pertaining to the electron transport through bipolar transistor with a non-uniform energy gap in the base region", *Solid-State Electronics*, Vol. 28, No. 11, pp. 1101-1103, 1985.
- [22] H. Tran, "Physics-based analytical modeling of SiGe heterojunction bipolar transistors for high-speed integrated circuits", Ph.D. thesis, University of Technology Dresden, Germany, 2006.
- [23] J.L. Moll and J.M. Ross, "The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity", *Proc. IRE*, Vol. 44, pp. 72-78, 1956.

- [24] J.J. Ebers and J.L. Moll, "Large-Signal Behavior of Junction Transistors", Proc. IRE, Vol. 42, pp. 1761-1772, 1954.
- [25] H.K. Gummel and H.C. Poon, "An Integral Charge-Control Model for Bipolar Transistors", BSTJ Vol. 49, pp. 827-852, 1970.
- [26] H.K. Gummel, "A Charge-Control Relation for Bipolar Transistors", BSTJ Vol. 49, pp. 115-120, 1970.
- [27] M. Friedrich and H.-M. Rein, "Analytical current-voltage relations for compact SiGe HBT models - Part I: The idealized HBT", IEEE Trans. Electron Dev., vol. 46, pp. 1384-1393, 1999.
- [28] M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors with special regard to high current densities and two-dimensional effects", Ph. D. thesis (in German), Ruhr-University Bochum, Bochum, Germany, 1988.
- [29] A. Joseph, J. Cressler, D. Richey, R. Jaeger, D. Hareme, "Neutral base recombination and its influence on the temperature dependence of Early voltage and current gain - Early voltage product in UHV/CVD SiGe heterojunction bipolar transistors", IEEE. Trans. Electron Dev., Vol. 44, pp. 404-412, 1997.
- [30] M. Peter, J. Slotboom, D. Terpstra, "Impact ionization and neutral base recombination in SiGe HBTs", Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), pp. 58-61, 1999.
- [31] M. Schröter, "HICUM status overview", HICUM Workshop, Bordeaux, France, May 2004.
- [32] S. Fregronese, T. Zimmer, C. Maneux, P. Sulima, "Barrier effects in SiGe HBTs: Modeling of high-injection base current injection", Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), pp. 104-107, 2004.
- [33] Q. Liang, J. Cressler, G. Niu, R. Murty, K. Newton, D. Hareme, "A physics-based transit time model applied to barrier effects", IEEE Trans. Electron Dev., Vol. 49, pp. 1807-1813, 2002.
- [34] E. Crabbe, J. Stork, G. Baccarani, M. Fischetti, S. Laux, "The impact of non-equilibrium transport on breakdown and transit time in bipolar transistors", Proc. IEDM, pp. 463-466, 1990.
- [35] J. Slotboom, G. Streutker, M. van Dort, H. Woerlee, A. Pruijmbloom, D. Gravesteijn, "Non-local impact ionization in silicon devices", Proc. IEDM, pp. 127-130, 1991.
- [36] G. Verzellesi, G. Baccarani, C. Canali, P. Pavan, L. Vendrame, E. Zanoni, "Prediction of impact-ionization-induced snap-back in ad-

- vanced Si-npn BJTs by means of a no-local analytical model for the avalanche multiplication factor", *IEEE Trans. Electron Dev.* Vol. 40, pp. 2296-2300, 1993.
- [37] L. I. Schiff, "Quantum Mechanics", McGraw-Hill Education; 3rd edition, 1968.
  - [38] S. Wang, "Fundamentals of semiconductor theory and device physics", Prentice Hall, NJ, ISBN 0-13-344409-0, 1989.
  - [39] C.-T. Sah, "Fundamentals of Solid-State Electronics", World Scientific, Singapore, 1991.
  - [40] E.O. Kane, "Theory of tunneling", *J. Appl. Phys.*, Vol. 32, No. 1, pp. 83-92, 1961.
  - [41] P. Hagelstein, S. Senturia, T. Orlando, "Introductory applied quantum and statistical mechanics", Wiley Interscience, NJ, 2004.
  - [42] S.M. Sze, "Physics of Semiconductor Devices", Wiley & Sons, New York, 1981.
  - [43] M. Schröter, H. Tran, W. Kraus, "Germanium profile design options for LEC-SiGe HBTs", *Solid-State Electronics*, Vol. 48, pp. 1133-1146, 2004.
  - [44] M. Schröter, Z. Yan, T.-Y. Lee, W. Shi, "A compact tunneling current collector breakdown model", *Proc. BCTM*, pp. 203-2-06, 1998.
  - [45] K. Lee, A.-H. Cho, K.-W. Park, B. Kim, "Improved VBIC model for SiGe HBTs with an unified model of heterojunction barrier effects", *IEEE Trans. Electron Dev.*, Vol. 53, No. 4, pp. 743-752, 2006.
  - [46] M. Schröter and T.-Y. Lee., "A physics-based minority charge and transit time model for bipolar transistors", *IEEE Trans. Electron Dev.*, vol. 46, pp. 288-300, 1999.
  - [47] D.E. Thomas und J.L. Moll, "Junction Transistor Short-Circuit Current Gain and Phase Determination", *Proc. IRE*, Vol. 46, pp. 1177-1184, 1958.
  - [48] J. Seitchik, A. Chatterjee und P. Yang, "An Accurate Bipolar Model for Large Signal Transient and AC Applications", *IEDM Tech. Dig.*, 1987, S. 244-247.
  - [49] M. Rudolph, "Limitations of current compact transit time models for III-V based HBTs", *Int. Microwave Symp. Digest*, pp. 487-490, 2008.
  - [50] J. Lindmayer und C.Y. Wrigley, "Fundamentals of Semiconductor Devices", D. Van Nostrand Comp. Inc., Toronto, 1965.

- [51] L. Chua and C.-C. Chang, "High-speed nonlinear circuit models for p-n junction diodes", *Int. J. Circ. Theory Applic.*, Vol. 16, pp. 157-190, 1988.
- [52] M. Sippilä, V. Porra, M. Valtonen, "Improved description of base dynamics in the modeling of bipolar transistors", *Int. J. Circuit Theory and Applic.*, Vol. 17, pp. 465-482, 1989.
- [53] J. Seitchik and J. Hamel, "Transient base dynamics of bipolar transistors in high injection", *IEEE Trans. Electron Dev.*, Vol. 41, pp. 2385-2390, 1994.
- [54] J. Parker, D. Roulston, J. Hamel, "Partitioned-charge-based BJT model using transient charge control relations for arbitrary doping and bias conditions" *IEEE Trans. Electron Dev.*, Vol. 30, No.4, pp. 605-612, 1993.
- [55] R.J. McDonald, "Generalized partitioned-charge-based bipolar transistor modeling methodology", *Electronics Letters*, Vol. 24, No. 21, pp. 1302-1304, Oct. 1988.
- [56] J. Hamel, C. Selvakumar, "The general transient charge control relation: a new charge control relation for semiconductor devices", *IEEE Trans. Electron Dev.*, Vol. 38, pp.1467-1476, 1991.
- [57] H. Klose und A. Wieder, "The Transient Integral Charge-Control Relation - a Novel Formulation of the Currents in a Bipolar Transistor", *IEEE Trans. Electron Dev.*, Vol. ED-34, pp. 1090-1099, 1987.
- [58] J. TeWinkel, "Extended Charge-Control Model for Bipolar Transistors", *IEEE Trans. Electron Dev.*, Vol. ED-20, pp. 389-394, 1973.
- [59] G. Hurkx, "A new approach to A.C. characterization of bipolar transistors", *Solid-State Electron.*, Vol. 31, pp. 1269-1275, 1988.
- [60] P.B. Weil und L.P. McNamee "Simulation of Excess Phase in Bipolar Transistors", *IEEE Trans. Circ. Syst.*, Vol. CAS-25, 1978, S. 114-116.
- [61] M. Schröter and H.-M. Rein, "Investigation of very fast and high-current transients in digital bipolar circuits by using a new compact model and a device simulator", *IEEE J. Solid-State Circuits*, Vol. 30, pp. 551-562, 1995.
- [62] J.G. Fossum und S. Veeraraghavan, "Partitioned Charge-Based Modeling of Bipolar Transistors for Non-Quasi-Static Circuit Simulation", *IEEE Electron Dev. Lett.*, Vol. EDL-7, pp. 652-654, 1986.

## **Chapter 5**

# **Geometry (Layout) Scaling**





In a compact model, the transistor  $T_i$  in Fig. 5.1 is approximated by an equivalent circuit consisting of a lumped element representation of the currents and charges discussed in the previous chapters. The equations contain an emitter area  $A_E$ , which is approximated by the emitter window area  $A_{E0} = b_{E0}l_{E0}$ . The equivalent circuit for the internal transistor then consists of  $T_i$  and an internal base impedance  $Z_{Bi}$  that connects the base terminal  $B'$  of  $T_i$  with the perimeter base node  $B^*$ .

In this chapter, the basic equations for the various physical effects and characteristics as function of geometry will be derived. Device simulation will be used to illustrate assumptions and to verify results. The main physical effects in both the 2D and 3D case are:

- Current crowding in the internal base region under the emitter, impacting the internal base resistance and perimeter injection.
- Carrier injection across the emitter perimeter with the associated current spreading and charge storage. Especially the transfer current and the minority charge are the most difficult variables to describe theoretically in an accurate way.
- Current spreading in the collector region, leading to a shift of the Kirk effect and related critical current towards higher current densities.
- Three-dimensional current flow and potential distribution in realistic structures. These effects are most pronounced and difficult (or at least complicated) to describe for the base and substrate region.

The numerical examples during the discussion are based on 2D device simulation results of the structures exhibited in Fig. 2.5b and Fig. 2.6b. The associated electrical characteristics are displayed in Fig. 5.2 and Fig. 5.3. In addition to the collector current also the normalized transconductance  $g_m/(I_C/V_T)$  has been inserted in order to provide a better look at non-ideal behavior. The latter can be observed for the BJT at lower current densities than for the HBT. The transit frequency curves look similar to the 1D case, except that the peak value is slightly lower. Further differences will be discussed in more detail in conjunction with the corresponding 2D effects.

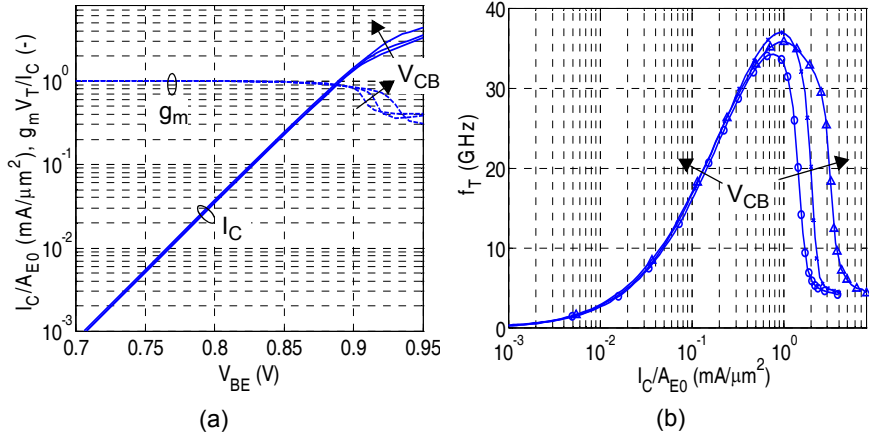


Fig. 5.2: Important electrical characteristics of the 2D BJT: (a) Collector current density and normalized transconductance. (b) Transit frequency.  
 $V_{BC}/V = 0.4$  (o), 0 (x), -2.3 ( $\Delta$ ).

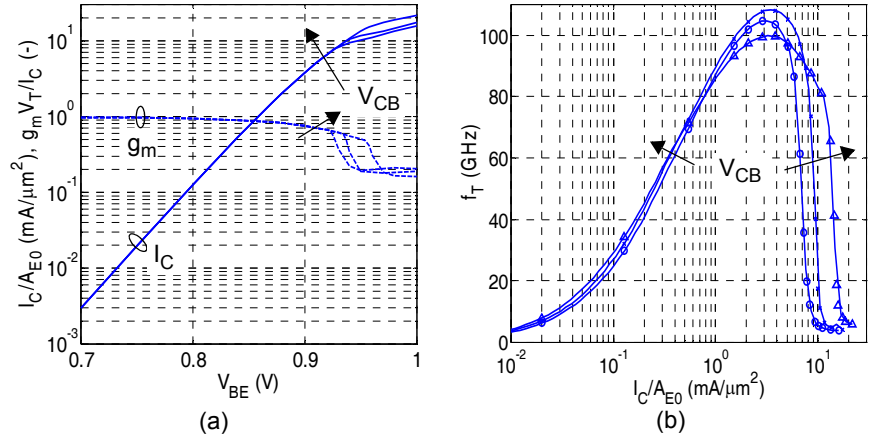


Fig. 5.3: Important electrical characteristics of the 2D HBT: (a) Collector current density and normalized transconductance. (b) Transit frequency.  
 $V_{BC}/V = 0.4$  (o), 0 (x), -2.3 ( $\Delta$ ).

## 5.2 Internal base resistance

The main purpose of this section is to provide an overview on (i) the physical mechanisms during DC, AC and transient operation that lead to the internal base resistance  $R_{Bi}$  or impedance  $Z_{Bi}$ , and (ii) the corresponding analytical treatment. Since  $R_{Bi}$  depends on bias, temperature, geometry and operation mode it is one of the most complicated elements in a bipolar transistor in terms of a theoretical description.

The top portion of Fig. 5.4a shows an enlarged view of the internal transistor for a structure with two long base stripes in parallel to the emitter (i.e. 2D case). Thus, from each side half of the internal base current  $i_{Bi}$  flows laterally into the base region, each supporting half of the internal transistor. Within an interval  $\Delta y$  a portion of the current is injected upwards into the emitter while the remaining current keeps flowing in lateral direction until the symmetry line at  $y = 0$ , where the terminal current has been completely injected into the emitter. Obviously, the lateral current flow causes a voltage drop in parallel to the junctions that leads to a larger forward bias at the edges  $y = [-b_{E0}/2, b_{E0}/2]$  compared with that at the center. For very high base currents, e.g. during transient operation, the voltage drop can lead to a significant decrease in bias of a portion of the internal transistor. The distributed current flow and its associated voltage drop can be described by a second-order differential equation. A closed-form solution is possible for certain special cases, which will be discussed below.

In principle, the problem could be solved numerically with a circuit simulator by using a multi-section transistor model as shown in the bottom portion of Fig. 5.4a. Here, the two symmetrical halves of the total transistor have been combined in the resistor and transistor elements. However, the complexity of such a multi-section model leads to a significant increase in computational effort, which is undesired for circuit design. The goal, therefore, is the representation of the distributed effects with a minimum number of elements, such as in the form of a lumped internal base resistance  $R_{Bi}$  or impedance  $Z_{Bi}$  and a lumped intrinsic transistor  $T_i$  that combines the discrete elements  $T_{l...n}$ . The resulting compact equivalent circuit is shown in Fig. 5.4b. Note, that  $v_{B'E'}$  corresponds to an effective internal voltage. The respective discussion in the subsequent sections is subdivided according to the different operating modes.



$$dr = r_{Sbi} \frac{dy}{l_{E0}}, \quad (5.1)$$

while the respective diode current is given by

$$I_D(y) = l_{E0} dy J_{Sbi} \exp\left(\frac{V(y)}{V_T}\right). \quad (5.2)$$

Here,  $J_{Sbi}$  is the saturation current density and  $V(y)$  is the voltage across the internal BE junction at location  $y$ . In order to obtain a lumped resistance element the current distribution  $I(y)$  and the voltage distribution  $V(y)$  need to be calculated.

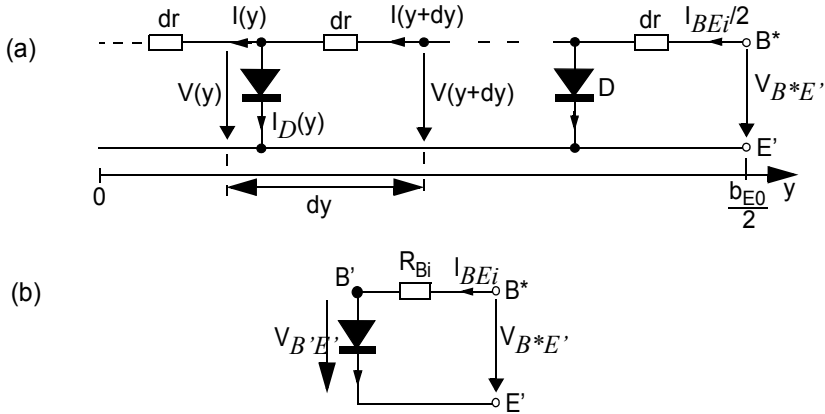


Fig. 5.5: For deriving the DC internal base resistance: (a) distributed resistor-diode network and (b) equivalent lumped representation.

In any spatial element  $dy$  the current balance yields

$$I(y + dy) - I(y) = I_D(y) \quad (5.3)$$

while the voltage loop gives

$$V(y + dy) - V(y) = dr I(y + dy). \quad (5.4)$$

Inserting  $I_D$  from (5.2) and  $dr$  from (5.1) allows after division by  $dy$  to write the differential equations

$$\frac{dI}{dy} = l_{E0} J_{Sbi} \exp\left(\frac{V(y)}{V_T}\right) \quad (5.5)$$

$$\frac{dV}{dy} = \frac{r_{SBi}}{l_{E0}} I. \quad (5.6)$$

Taking the second derivative of the current and inserting (5.6) gives a second-order differential equation in terms of only the current,

$$\frac{d^2 I}{dy^2} = \frac{l_{E0} J_{SBi}}{V_T} \exp\left(\frac{V(y)}{V_T}\right) \frac{dV}{dy} = \frac{1}{V_T} \frac{r_{SBi}}{l_{E0}} I \frac{dI}{dy} \quad (5.7)$$

where  $I_D/dy = dI/dy$  has been used. Realizing that

$$\frac{dI^2}{dy} = 2I \frac{dI}{dy} \quad (5.8)$$

allows to write

$$\frac{d^2 I}{dy^2} - \frac{1}{2V_T} \frac{r_{SBi}}{l_{E0}} \frac{dI^2}{dy} = 0, \quad (5.9)$$

which yields after integrating once

$$\frac{dI}{dy} - \frac{1}{2V_T} \frac{r_{SBi}}{l_{E0}} I^2 + C = 0 \quad (5.10)$$

with the integration constant  $C$ . The general solution of this still nonlinear differential equation is (e.g. [1])

$$I(y) = K_2 \tan\left(-\frac{K_1 + Cy}{K_2}\right) \quad (5.11)$$

$$\text{with } K_2 = \sqrt{(-C) \frac{2V_T l_{E0}}{r_{SBi}}}. \quad (5.12)$$

From the boundary condition  $I(y=0) = 0$  follows  $K_1 = 0$ . Furthermore, inserting the boundary condition  $I(b_{E0}/2) = I_{BEi}/2$  leads to

$$\frac{I_{BEi}}{2} = K_2 \tan\left(-\frac{C}{K_2} \frac{b_{E0}}{2}\right) = K_2 \tan\left(K_2 \frac{1}{2V_T} \frac{r_{SBi}}{l_{E0}} \frac{b_{E0}}{2}\right) \quad (5.13)$$

where (5.12) has been used. Defining the variable

$$Z = K_2 \frac{r_{SBi} b_{E0}}{V_T 4l_{E0}} \quad (5.14)$$

yields the transcendental equation

$$Z \tan(Z) = \eta = \frac{r_{SBi} b_{E0}}{V_T 8l_{E0}} I_{BEi} \quad (5.15)$$

from which  $Z$  and, thus,  $K_2$  can be determined numerically. The final solution for the current distribution is obtained after inserting  $K_2$  and  $K_I$  into (5.11) and using (5.12) and (5.14):

$$I(y) = \frac{4l_{E0} V_T}{r_{SBi} b_{E0}} Z \tan\left(Z \frac{y}{b_{E0}/2}\right). \quad (5.16)$$

The voltage distribution follows from inserting above equation into (5.6) and after integration from the arbitrary location  $y$  to the emitter edge,

$$V(y) = V\left(\frac{b_{E0}}{2}\right) - 2V_T \ln\left(\frac{\cos\left(Z \frac{2y}{b_{E0}}\right)}{\cos(Z)}\right). \quad (5.17)$$

The obtained solutions for the voltage distribution are compared in Fig. 5.6 with the results from device simulation. Excellent agreement is obtained up to current densities somewhat above peak  $f_T$  for wide (1  $\mu\text{m}$ ) emitter. The deviations observed at high current densities are caused by the bias and, hence, spatial dependence of  $r_{SBi}$ , (cf. next section). If  $r_{SBi}$  becomes significantly dependent on  $y$  an additional term needs to be added to the differential equation (5.9). However, it is questionable whether a general solution can then still be obtained. Also, there is little practical relevance of such an extension for such high current densities. Note, that inserting a spatial average of  $r_{SBi}$  is not sufficient to eliminate the deviation observed in Fig. 5.6. In a structure with narrow emitter (e.g. 0.2  $\mu\text{m}$ ) the lateral voltage drop is barely visible up to very high current densities.

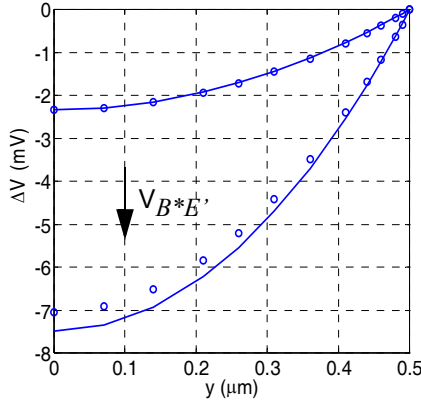


Fig. 5.6: Lateral voltage drop  $\Delta V = V(y) - V(b_{E0}/2)$  in the neutral base region (at  $x = 0.04 \mu\text{m}$ ) obtained from device simulation (symbols) and from theory (solid lines) for a wide ( $b_{E0} = 1 \mu\text{m}$ ) CED HBT structure operated at peak  $f_T$  ( $V_{B^*E'} = 0.89 \text{V}$ ) and at a high current density ( $f_T/2$ ,  $V_{B^*E'} = 0.925 \text{V}$ ). Note that only half of the structure is displayed and a spatial average of  $r_{SBi}$  was inserted for the calculation.

The next step is to derive a lumped resistance element that produces the same terminal characteristic  $I_{BEi}(V_{B^*E'})$  as the distributed equivalent circuit. From Fig. 5.5b follows that the lumped resistance is defined as

$$R_{Bi} = \frac{V_{B^*E'} - V_{B'E'}}{I_{BEi}} \quad (5.18)$$

with  $V_{B^*E'} = V(b_{E0}/2)$ . Integrating the injection current contributions  $I_D$  over the total emitter width yields

$$I_{BEi} = 2 \int_{I_D(0)}^{I_D(b_{E0}/2)} dI = I_{SBi} \exp\left(\frac{V_{B^*E'}}{V_T}\right) \cos(Z) \frac{\sin(Z)}{Z} \quad (5.19)$$

with  $dI$  from (5.5), and  $V(y)$  from (5.17). On the other hand, the internal base current is also given from the equivalent circuit in Fig. 5.5b as

$$I_{BEi} = I_{SBi} \exp\left(\frac{V_{B'E'}}{V_T}\right) = I_{SBi} \exp\left(\frac{V_{B^*E'} - R_{Bi} I_{BEi}}{V_T}\right). \quad (5.20)$$

From this and (5.19) the lumped internal base resistance follows as



$$R_{Bi} = \frac{V_T}{I_{BEi}} \ln \left[ \frac{Z}{\cos(Z) \sin(Z)} \right]. \quad (5.21)$$

To find the resistance at negligible current crowding, the  $\ln[\ ]$  term is expanded into a series,

$$\ln[\dots] \cong \ln \left[ \frac{Z}{\left(1 - \frac{Z^2}{2}\right) \left(Z - \frac{Z^3}{6}\right)} \right] \cong -\ln \left[ 1 - \frac{2Z^2}{3} \right] \cong \frac{2Z^2}{3}. \quad (5.22)$$

Since  $Z$  depends on  $I_{BEi}$ , series expansion of (5.15) permits to express  $Z$  by  $I_{BEi}$ :

$$Z^2 \cong \eta = I_{BEi} \left( \frac{r_{SBi}}{V_T} \frac{b_{E0}}{8l_{E0}} \right). \quad (5.23)$$

Inserting this result together with (5.22) into (5.21) yields the internal base resistance at negligible current crowding,

$$R_{Bi,nc} = r_{SBi} \frac{b_{E0}}{12l_{E0}}, \quad (5.24)$$

where the sheet resistance at low injection has to be inserted. This result was first derived in [2]. The bias dependence of  $r_{SBi}$  will be discussed later in more detail.

It is convenient to normalize  $R_{Bi}$  in (5.21) to  $R_{Bi,nc}$ , which is accomplished by extension with (5.24) and recognizing (5.15):

$$R_{Bi} = R_{Bi,nc} \psi(\eta). \quad (5.25)$$

The current-crowding factor  $\eta$  is defined by (5.15). The *emitter current-crowding function*

$$\psi(\eta) = \frac{3}{2\eta} \ln \left[ \frac{Z}{\cos(Z) \sin(Z)} \right] \quad (5.26)$$

was first derived in [3]. Figure 5.7 shows  $\psi(\eta)$  with  $Z$  from (5.15) over the practically relevant range of  $\eta$ . Note that  $2\eta/3 = R_{Bi,nc} I_{BEi} / V_T$  corresponds to a normalized voltage drop across the internal base resistance  $R_{Bi,nc}$ .

The obtained result for  $\psi(\eta)$  is based on the requirement that the characteristic  $I_{BEi}(V_{B^*E'})$  of the lumped model equals that of the distributed representation. This definition makes sense from a circuit design and parameter determination point of view since it ensures a correct *terminal characteristic*. However, the first calculation of emitter current crowding in [4] was based on the equivalence of *power dissipation*, yielding

$$\psi_p(\eta) = 3 \frac{\tan(Z) - Z}{Z \tan^2(Z)}. \quad (5.27)$$

Fundamentally, this result should not be used in compact models since it does not give the correct terminal characteristic. The deviation is negligible though at sufficiently low currents as can be observed in Fig. 5.7.

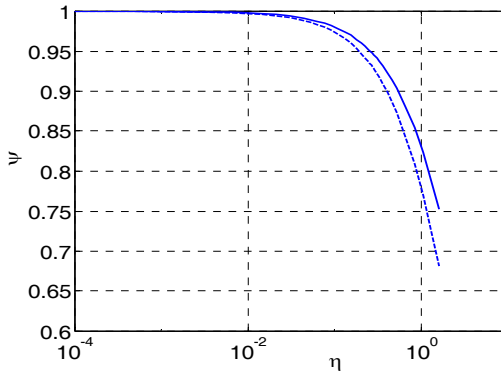


Fig. 5.7: Emitter current crowding function vs. normalized voltage drop (crowding factor): comparison between  $\psi$  (solid line) and  $\psi_p$  (dashed line), both with the corresponding bias dependent value for  $r_{SBi}$  (spatial average).

### 5.2.1.2 Conductivity modulation

The internal base sheet resistance is defined as

$$r_{SBi} = \left( q \int_{x_1}^{x_2} \mu_p p dx \right)^{-1}, \quad (5.28)$$

where the integration is performed along a line underneath the emitter window. The integration extends over the region  $[x_1, x_2]$  that represents the cross-section through which the lateral base current flows. As Fig. 5.8

shows for different technologies,  $r_{SBi}$  depends more or less strongly on the bias condition. At low injection,  $r_{SBi}$  already drops significantly in BJTs while it only varies by about 15% in HBTs. At higher injection a sharp drop is observed due to the onset of the high-current effect and the associated strong increase of the hole density and its vertical spread.

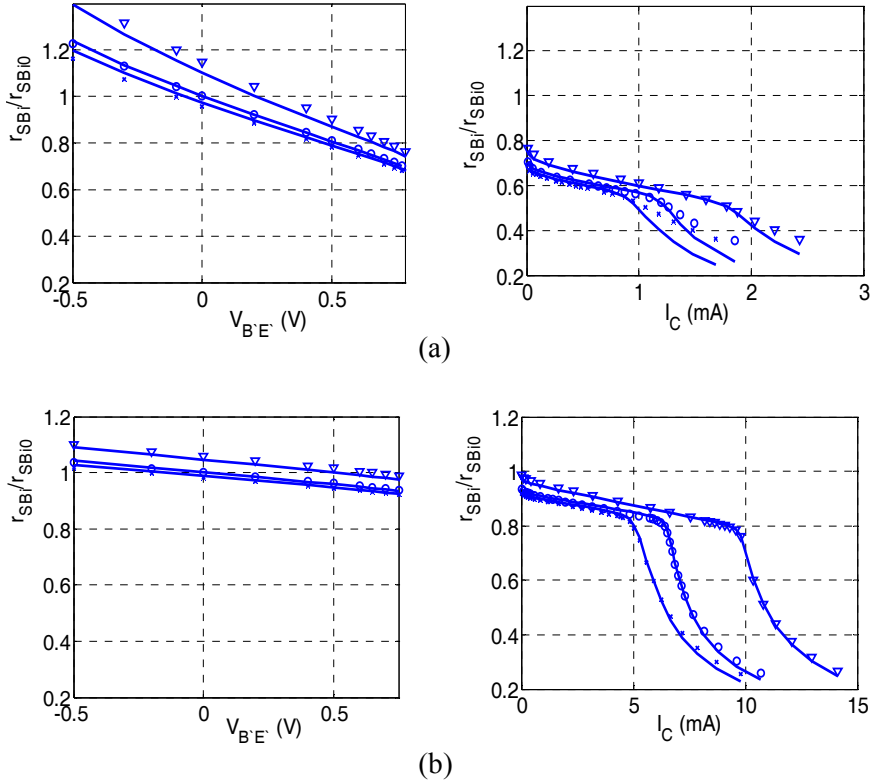


Fig. 5.8: Bias dependence of normalized internal base sheet resistance at different voltages for (a) BJT and (b) CED HBT: comparison between 1D device simulation (symbols) and analytical model (5.31).  $Q_{p0}/\text{fF} = 16.9$  (BJT), 79 (HBT) was determined at  $V_{B'E'} = 0.8\text{V}$  (BJT, HBT).  $V_{B'C'}/\text{V} = -2.3$  ( $\nabla$ ), 0 ( $\circ$ ), 0.4 ( $\times$ ). The actual zero-bias charge is  $Q_{p0}/\text{fF} = 16.2$  (BJT), 78.6 (HBT).

Fundamentally, there are three factors that contribute to a bias dependence of  $r_{SBi}$ : (i) field and doping dependent hole mobility; (ii) injection dependent hole density; (iii) bias dependent current flow cross-section (i.e. integration limits). These effects are differently pronounced in different

types of bipolar technologies. In older BJT technologies with a moderately doped base a variation of the junction voltages can cause a significant change in (neutral) base width and, hence, in conductivity even at low injection. At high current densities, the injection of holes into the collector leads to a widening of the effective cross-section for the lateral hole current and a fairly sudden drop of  $r_{SBi}$  (at approximately the critical current density  $J_{CK}$ ). If at very high current densities the injection of the electrons approaches the base doping concentration, the hole density starts to increase, which causes  $r_{SBi}$  to drop even further. These effects, also known as conductivity modulation, are much less pronounced in advanced SiGe HBTs due to their highly doped base and the valence band barriers which prevent the spreading of holes into the adjacent collector and emitter regions.

If the integration limits in (5.28) change with bias, the doping dependence of the mobility comes into play, since it acts as a weight factor for the hole density. For HBTs and advanced BJTs, the fairly symmetrical base doping profile causes  $\mu_p$  to increase for neutral base region expansions in both junction directions (due to the drop of the doping concentration there), leading to a relative increase of related hole contributions. This results in an amplification of the corresponding change. At high current densities the barrier at the BC junction causes  $r_{SBi}$  to decrease rapidly once the injection zone starts to form. The influence of the Ge variation is negligible. For older BJT generations, the base doping peak is defined by the emitter doping, so that the total doping increases towards the BE junction. This results in a drop of  $\mu_p$  and, thus, a suppression of the corresponding contribution to the integral.

For all transistor types, the field dependence of  $\mu_p$  is negligible unless the gradient of the hole quasi-fermi potential increases significantly. This practically never happens under DC or low-frequency operation, but may occur during high-speed switching when the dynamic base current can become very large and causes a large lateral voltage drop. Transient operation will be discussed later though.

According to Fig. 5.8 the bias dependence of  $r_{SBi}$  needs to be modelled to make a compact model generally applicable. For a given bias point, it is always possible to define a spatially averaged hole mobility in (5.28) and to write

$$r_{Sbi} = \left( q \bar{\mu}_p \int_{x_1}^{x_2} p dx \right)^{-1} = \frac{1}{\bar{\mu}_p \bar{Q}_{pr}}. \quad (5.29)$$

Here,  $\bar{Q}_{pr}$  is the hole charge density within the cross-section of the lateral hole current flow. For further discussion, it is convenient to normalize the sheet resistance to its zero-bias value, yielding

$$\frac{r_{Sbi}}{r_{Sbi0}} = \frac{\bar{\mu}_{p0} Q_{p0}}{\bar{\mu}_{p0} Q_{p0} + \bar{\mu}_{pe} \Delta Q_{pe} + \bar{\mu}_{pc} \Delta Q_{pc}}, \quad (5.30)$$

where the charge densities were replaced by the corresponding actual charges. Now the bias dependent integral in the denominator has been split into a zero-bias component  $Q_{p0}$  and bias dependent contributions located at the emitter and collector side of the neutral base.  $\Delta Q_{pe}$  and  $\Delta Q_{pc}$  are defined w.r.t. zero-bias and, thus, disappear at zero bias. Furthermore,  $\bar{\mu}_{pe}$  and  $\bar{\mu}_{pc}$  are the associated mobilities in the extended base regions. As discussed earlier, an extension of the neutral base boundaries causes  $\bar{\mu}_{pc}$  to increase in all transistor types, while  $\bar{\mu}_{pe}$  increases in HBTs and decreases in older BJTs. However, for typical operating conditions with a forward biased BE junction and reverse biased BC junction, the charges have an opposite sign. Therefore, the bias dependent terms in (5.30) generally compensate each other partially. In older BJTs this compensation is more pronounced than in HBTs.

For compact modeling, (5.30) is not suitable since the average mobilities are unknown. With  $\Delta Q_p$  as actual charge variation determined from small-signal parameters and defining  $\bar{\mu}_{pQ} = (\bar{\mu}_{pe} \Delta Q_{pe} + \bar{\mu}_{pc} \Delta Q_{pc}) / \Delta Q_p$  allows to write the approximation

$$\frac{r_{Sbi}}{r_{Sbi0}} \approx \frac{Q_{p0r}}{Q_{p0r} + \Delta Q_p} \quad \text{with} \quad Q_{p0r} = \frac{\bar{\mu}_{p0}}{\bar{\mu}_{pQ}} Q_{p0}. \quad (5.31)$$

The comparison of above expression with device simulation in Fig. 5.8 shows excellent agreement for the HBT over the whole bias region and for the BJT in the practical relevant bias range, if  $Q_{p0r}$  is determined at forward bias (around  $V_{B'E'} = 0.8V$  here). The observed deviations in the current dependence at very high injection and in the  $V_{B'C'}$  dependence are caused by the bias dependence of the average mobility  $\bar{\mu}_{pQ}$ .

Since in older BJTs the impact of  $\mu_p$  on  $r_{SBI}$  is often negligible for a simultaneous change of  $V_{B'E'} = V_{B'C'}$ , (5.31) was used in [5, 6] for the determination of  $Q_{p0}$  from sheet resistance measurement. As it turns out this condition still works quite well for HBTs. Best results are obtained either at reverse bias conditions or within a fairly narrow forward bias region (around 0.8V here). The determination from data at constant  $V_{B'C'}$  yields best results around  $V_{B'C'} = 0$  and at forward bias (around 0.8V here).

According to (5.28) the sheet resistance is defined along a vertical line under the emitter. As a consequence of its bias dependence,  $r_{SBI}$  can become a function of lateral dimension during current-crowding. For the 2D case considered so far the BE voltage increases towards the perimeter edge of the emitter, leading to a lower value of  $r_{SBI}$  there. This in turn reduces the corresponding resistance and, hence, the DC current crowding effect. Figure 5.9 displays the lateral variation of  $r_{SBI}$  obtained from device simulation for a transistor with a wide (1  $\mu\text{m}$ ) emitter, operated at around peak  $f_T$  and in the high current range. As can be seen, even for such a wide emitter the lateral variation of  $r_{SBI}$  is quite small and does not exceed 4%. For comparison, also the corresponding normalized vertical electron current density has been inserted, which exhibits the different levels of current crowding quite clearly due to the exponential dependence on the voltage drop.

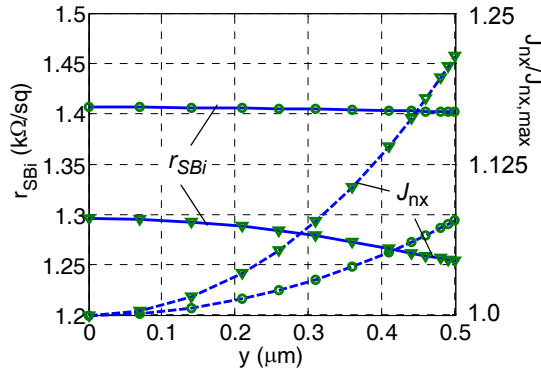


Fig. 5.9: Lateral spatial dependence of the internal base sheet resistance (solid lines with symbols) from device simulation of a wide emitter ( $b_{E0} = 1 \mu\text{m}$ ) CED HBT for two different bias points at  $f_{T,peak}$  ( $V_{B'E'} = 0.89\text{V}$ , circles) and at  $f_{T,peak}/2$  ( $V_{B'E'} = 0.925\text{V}$ , triangles). For comparison, also the corresponding normalized vertical electron current density  $J_{nx}$  has been added (dashed lines).

### 5.2.2 Small-signal operation

For the small-signal case the diode elements in Fig. 5.5 can be replaced by their associated conductance  $dG$  and capacitance  $dC$ :

$$dG = \bar{G}l_{E0}dy, \quad dC = \bar{C}l_{E0}dy. \quad (5.32)$$

$\bar{G}$  and  $\bar{C}$  represent the corresponding area specific values *at the given bias point*. They will be discussed later in more detail. Defining the admittance

$$d\mathbf{Y} = dG + j\omega dC = l_{E0}(\bar{G} + j\omega\bar{C})dy = l_{E0}\mathbf{Y}dy \quad (5.33)$$

then leads to the equivalent circuit shown in Fig. 5.10a where  $dr$  is given by (5.1). Within the spatial element  $dy$  the current balance now yields in frequency domain

$$I(y+dy) - I(y) = dI_D(y) = dy l_{E0} \mathbf{Y} V(y) \quad (5.34)$$

while the voltage loop gives

$$V(y+dy) - V(y) = dr I(y+dy). \quad (5.35)$$

The combination of these two equations yields after inserting (5.1) the linear second-order differential equation

$$d^2 I / dy^2 = \mathbf{Y} r_{SBi} I. \quad (5.36)$$

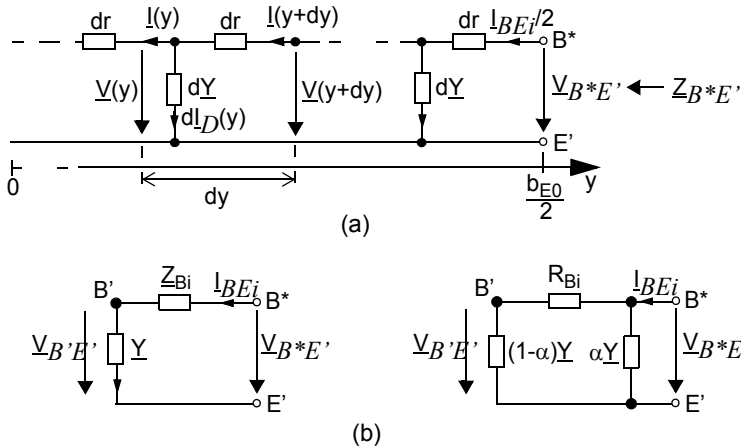


Fig. 5.10: Small-signal dynamic emitter current crowding: (a) distributed representation and (b) possible lumped representations.

Assuming that  $\bar{G}$ ,  $\bar{C}$  and  $dr$  do not depend on  $y$  (i.e. neglecting possible DC current crowding), the general solution reads

$$I(y) = \underline{A} \cosh(\underline{k}y) + \underline{B} \sinh(\underline{k}y) \quad (5.37)$$

with

$$\underline{k} = \sqrt{\underline{Y} r_{SBi}}. \quad (5.38)$$

From the symmetry condition,  $I(y) = -I(-y)$ , follows  $\underline{A} = 0$ , while the condition  $I(b_{E0}/2) = I_{BEi}/2$  at the terminal yields the constant

$$\underline{B} = \frac{I_{BEi}}{2 \sinh(\underline{k}b_{E0}/2)}. \quad (5.39)$$

Inserting  $\underline{A}$  and  $\underline{B}$  into (5.37) gives the solution

$$I(y) = \frac{I_{BEi}}{2} \frac{\sinh(\underline{k}y)}{\sinh(\underline{k}b_{E0}/2)} \quad (5.40)$$

and

$$V(y) = \frac{I_{BEi}}{2l_{E0}\underline{Y}} \underline{k} \frac{\cosh(\underline{k}y)}{\sinh(\underline{k}b_{E0}/2)}. \quad (5.41)$$

This solution agrees excellently with device simulation results under the assumptions made.

Defining the lumped junction admittance

$$\underline{Y} = \underline{Y}b_{E0}l_{E0} = G + j\omega C \quad (5.42)$$

and inserting this along with  $V(b_{E0}/2) = V_{B^*E'}$ , from (5.41) yields a closed-form solution for the internal input impedance

$$\underline{Z}_{B^*E'} = \frac{V_{B^*E'}}{I_{BEi}} = \frac{1}{\underline{Y}} \frac{\underline{k}b_{E0}}{2} \coth\left(\underline{k} \frac{b_{E0}}{2}\right), \quad (5.43)$$

which was first derived in [2]. Figure 5.11 shows the frequency dependence of the normalized input impedance of the structure in Fig. 5.4a for a bias point at peak  $f_T$ . The analytical solution above agrees excellently with the device simulation results over the whole frequency range despite slight DC current crowding. At low frequencies  $G \gg \omega C$ , so that  $\text{Re}\{\underline{Z}_{B^*E'}\} \approx 1/G$  and  $\text{Im}\{\underline{Z}_{B^*E'}\} \approx -\omega C/G^2$ . Once  $\omega C$  reaches  $G$  at higher frequencies the imaginary part peaks and the real part starts to drop rapidly



towards  $R_{Bi,nc}$ . The latter is reached at about  $f_T/5$ . In this medium frequency range,  $\text{Im}\{\underline{Z}_{B^*E'}\} \approx 1/\omega C$ . At very high frequencies the capacitive current across the junction at the emitter edge ( $y = b_{E0}/2$ ) starts to shunt the regions towards the center of the emitter ( $y = 0$ ). Thus, both  $\text{Re}\{\underline{Z}_{B^*E'}\}$  and  $\text{Im}\{\underline{Z}_{B^*E'}\}$  drop further with a decreasing slope though, that is determined by the portion of the emitter region carrying most of the AC current (cf. emitter utilization factor further below). This effect is sometimes called *lateral non-quasi-static effect*.

The goal is to obtain a simple lumped equivalent circuit such as those shown in Fig. 5.10b. For the one on the left hand side, the lumped internal base impedance follows directly from (5.43) and is given by

$$\underline{Z}_{Bi} = \frac{V_{B^*E'} - V_{B'E}}{I_{BEi}} = \frac{1}{\underline{Y}} \left[ \frac{k b_{E0}}{2} \coth\left(k \frac{b_{E0}}{2}\right) - 1 \right]. \quad (5.44)$$

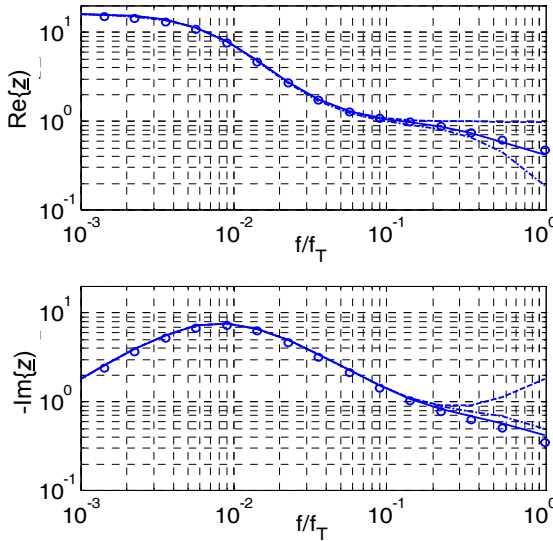


Fig. 5.11: Frequency dependent normalized internal input impedance,  $\underline{Z} = \underline{Z}_{B^*E'}/R_{Bi,nc}$ , vs. normalized frequency  $f/f_T$ :  $\underline{Z}_{B^*E'}$  from (5.43) (solid line),  $\underline{Z}_{B^*E'}/f_T$  from (5.49) (dashed line), device simulation as reference (circles), and  $\Pi$ -EC with fitting parameters  $\alpha = 0.2$  and  $R_{Bi} = 175\Omega$  (dash-dotted line). The calculations are for a wide emitter ( $b_{E0} = 1\mu\text{m}$ ) CED HBT structure.

For further discussion, it is convenient to express the argument of the coth directly by the low-current DC base resistance (using (5.24)) and the total junction admittance (using (5.42)), which leads to the *small-signal current crowding factor*

$$\eta_{ac}(\omega) = \frac{k b_{E0}}{2} = \sqrt{3 \underline{Y} R_{Bi, nc}}. \quad (5.45)$$

Inserting into (5.44) allows to write

$$\underline{Z}_{Bi} = R_{Bi, nc} \underline{\psi}_f(\eta_{ac}) \quad (5.46)$$

with the *AC current-crowding function*

$$\underline{\psi}_{ac}(\eta_{ac}) = 3 \frac{\eta_{ac} \coth(\eta_{ac}) - 1}{\eta_{ac}^2}. \quad (5.47)$$

In order to obtain a better understanding of the dependence of  $\underline{Z}_{B^*E}$  on frequency and to be able to convert the exact solution into a simple equivalent circuit such as those shown in Fig. 5.10b, higher order frequency terms are discarded. Assuming  $\eta_{ac} \ll 1$  allows a series expansion of (5.43), yielding

$$\underline{Z}_{B^*E} = \frac{1}{\underline{Y}} \eta_{ac} \left( \frac{1}{\eta_{ac}} + \frac{\eta_{ac}}{3} - \frac{\eta_{ac}^3}{45} + \dots \right) \approx \frac{1}{\underline{Y}} \left( 1 + \frac{\eta_{ac}^2}{3} - \frac{\eta_{ac}^4}{45} \right) \quad (5.48)$$

after dropping *frequency* terms higher than second-order. Inserting (5.45) for  $\eta_{ac}$  gives the *low-frequency* internal input impedance

$$\underline{Z}_{B^*E, lf} = \frac{1}{\underline{Y}} + R_{Bi, nc} - \frac{\underline{Y} R_{Bi, nc}^2}{5}. \quad (5.49)$$

The frequency dependence is shown for comparison in Fig. 5.11. The agreement is quite good up to about  $f_T/5$  and acceptable up to  $f_T/5$ . Also shown is the result obtained for the  $\Pi$ -EC in Fig. 5.10b (left) after fitting proper values for the variables  $\alpha$  and  $R_{Bi}$ .

The low-frequency internal base impedance follows from (5.49) after subtracting  $1/\underline{Y}$  as

$$\underline{Z}_{Bi,lf} = R_{Bi,nc} \left[ 1 - \frac{\underline{Y}R_{Bi,nc}}{5} \right] = R_{Bi,nc} \underline{\Psi}_{ac,lf}(\underline{\eta}_{ac}). \quad (5.50)$$

Above approximation is justified for the real part since  $GR_{Bi,nc}/5 = I_{BEI}R_{Bi,nc}/(5V_T) \ll 1$  as per the assumption of negligible DC current crowding. Note that a significantly better approximation of the exact solution can be obtained by converting the term in the brackets to the denominator,

$$\underline{Z}_{Bi,lf} = R_{Bi,nc} \left[ 1 + \frac{\underline{Y}R_{Bi,nc}}{5} \right]^{-1}, \quad (5.51)$$

which leads to kind of an error compensation. This solution can also be converted directly into an equivalent circuit [7]. Finally, the small-signal *zero-frequency* expression,

$$Z_{Bi0} = R_{Bi,nc} \left[ 1 - \frac{GR_{Bi,nc}}{5} \right] \cong R_{Bi,nc}, \quad (5.52)$$

leads to the same result as the DC solution.

The AC current crowding function  $\underline{\Psi}_{ac}(\underline{\eta}_{ac})$  determines the drop of the base impedance with frequency from its low-frequency value and is a measure for the severity of the lateral NQS effect. Figure 5.12 exhibits both  $\underline{\Psi}_{ac}$  and  $\underline{\eta}_{ac}$  as a function of frequency and emitter width.  $\text{Re}\{\underline{\eta}_{ac}\}$  is constant as long as  $G \gg \omega C$ , while  $\text{Im}\{\underline{\eta}_{ac}\}$  always increases with frequency (Fig. (a)).

At low frequencies,  $\text{Re}\{\underline{\Psi}_{ac}\} \approx \Psi_{ac}$  since  $\text{Im}\{\underline{\Psi}_{ac}\}$  is very small (Fig. 5.12b). At high frequencies, i.e. for  $\omega C \gg G$ , the argument of the square root in (5.45) is imaginary, so that the real part of  $\underline{\Psi}_{ac}$  equals the imaginary part. At these frequencies  $\underline{Y} \approx \omega C = \omega(f/f_T)g_m$  for the common-emitter operation considered here. For the same  $f/f_T$ , therefore,  $\underline{\eta}_{ac}$  increases with bias via  $g_m$ . Thus, at higher bias  $\text{Re}\{\underline{\Psi}_{ac}\}$  and, hence,  $\underline{\Psi}_{ac}$  starts to decrease already at lower frequency as observed in Fig. 5.12b.  $\text{Im}\{\underline{\Psi}_{ac}\}$  also increases first with  $f$  but then drops according to  $1/\underline{\eta}_{ac}$ .

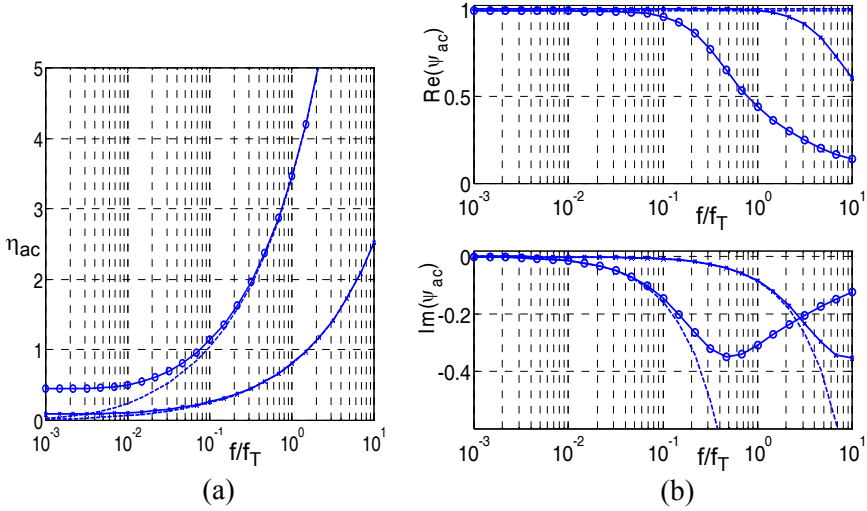


Fig. 5.12: (a) AC emitter current crowding factor  $\text{Re}\{\eta_{ac}\}$  (solid lines with symbols) and  $\text{Im}\{\eta_{ac}\}$  (dashed lines) vs. normalized frequency for the bias point  $V_{B^*E'} = 0.8V$  at  $\approx f_{T,\text{peak}}/4$  (crosses) and for  $V_{B^*E'} = 0.89V$  at  $f_{T,\text{peak}}$  (circles) of the internal CED HBT structure ( $b_{E0} = 1\mu\text{m}$ ). (b) AC emitter current crowding function  $\text{Re}\{\psi_{ac}\}$  (top) and  $\text{Im}\{\psi_{ac}\}$  (bottom) vs. normalized frequency for the same bias conditions as in (a): comparison between exact solution (5.47) (solid lines with symbols) and low-frequency approximation (5.50) (dashed lines).

Unfortunately, it has been impossible so far to find a closed-form analytical solution for the general case of DC *and* AC current crowding. Numerical solutions for such cases are presented in [3][8][9][10]. In [8] a reduced emitter width  $b_{E,dc}$  is calculated from DC crowding that represents the “effective” width for the DC current injection. Inserting  $b_{E,dc}$  into the AC solution derived above yields fairly good agreement for the spatial ( $y$ ) dependence of the injected AC current from low to high frequencies. For non-negligible DC current crowding the small-signal zero-frequency resistance needs to include the bias dependence of  $R_{Bi}$  and is given by (cf. (5.18))

$$R_{bi0} = \frac{d(V_{B^*E'} - V_{B'E'})}{dI_{BEi}} = R_{Bi} + I_{BEi} \frac{dR_{Bi}}{dI_{BEi}}. \quad (5.53)$$

This result can also be obtained from a circuit simulator representation.

In microwave applications often a so-called *emitter utilization factor* is used that is a measure for the AC current crowding. The factor is defined w.r.t. the injected current density from the emitter as (e.g. [11])

$$F_{EU} = \left| \left( 2l_{E0} \int_0^{b_{E0}/2} J_{nx}(y) dy \right) / (b_{E0} l_{E0} J_{nx}|_{b_{E0}/2}) \right| \quad (5.54)$$

and basically represents the fraction of the emitter width (or area) that carries most of the injected current. Note that only the current component of the internal transistor (cf. Fig. 5.1) is considered. The emitter current density is related to the base current density distribution through (5.34)

$$J_{nx}(y) = \beta \frac{dI_D}{dy l_{E0}} = \beta \bar{Y} V(y). \quad (5.55)$$

Since AC current crowding only occurs for sufficiently high frequencies  $f > f_\beta$  (cf. Fig. 5.12b), the small-signal current gain can be approximated by  $\beta \approx -j f_T / f$ . The transit frequency  $f_T$  as well as  $\bar{Y}$  depend on the small-signal variables at each location  $y$  which, however, are spatially independent according to the assumption of negligible DC current crowding. Therefore,  $\beta$  and  $\bar{Y}$  cancel in (5.54) so that  $F_{EU}$  can be evaluated from inserting (5.41)

$$F_{EU} = \left| \frac{2 \int_0^{b_{E0}/2} \cosh(\underline{\kappa} y) dy}{b_{E0} \cosh(\underline{\kappa})} \right| = \left| \frac{1}{\underline{\kappa}} \tanh(\underline{\kappa}) \right|. \quad (5.56)$$

with the argument  $\underline{\kappa} = \underline{\kappa} b_{E0} / 2$ .

Figure 5.13 shows  $F_{EU}$  as a function of emitter width for selected frequencies up to  $1.5f_T$  and also as a function of frequency for selected widths. Note that a frequency of  $1.5f_T$  can be relevant for harmonic distortion. The results show that in a  $0.2\mu\text{m}$  wide device AC current crowding is negligible even for  $1.5f_T$  while it is severe even at  $0.5f_T$  in a wide device that is often used in, e.g., transistors for power amplifiers.

In the derivation above, the elements  $G$  and  $C$  in (5.42) were not specified in detail. For a dynamic BC short, one obtains

$$G = g_{BEi} \quad \text{and} \quad C = C_{jEi} + C_{dEi}, \quad (5.57)$$

while for a dynamic CE short

$$G = g_{BEi} + g_{BCi} \quad \text{and} \quad C = C_{jEi} + C_{dE} + C_{jCi} + C_{dC}. \quad (5.58)$$

Both are ideal cases and may not be exactly encountered in practice. In typical applications, one will generally find  $C_{jCi} + C_{dC} \ll C_{jEi} + C_{dE}$ . However, if the circuit causes a significant Miller effect, the BC voltage change can become much larger than the BE voltage change, so that the AC current through  $C_{jCi} + C_{dC}$  may not be negligible any more. It obviously depends on the application which elements need to be taken into account in  $G$  and  $C$ , and there is no “correct” or “incorrect” solution at the transistor level theory.

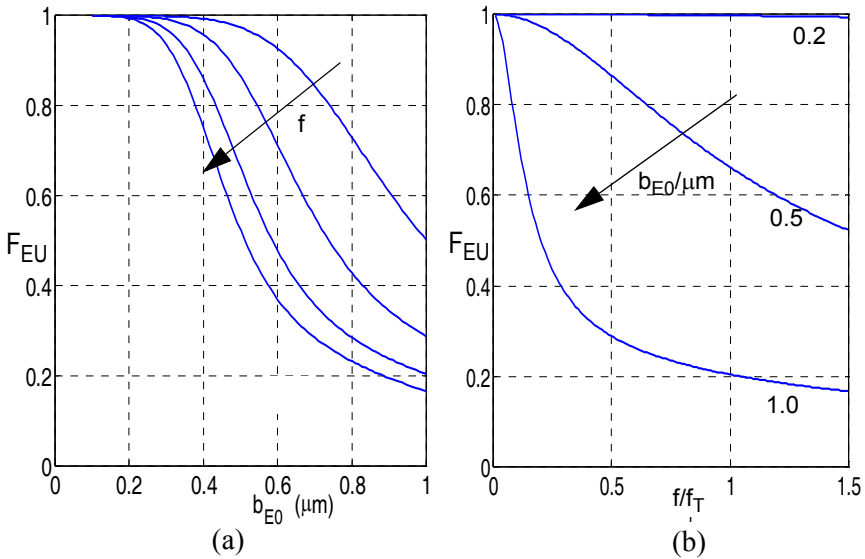


Fig. 5.13: Emitter utilization factor calculated from (5.56) with the parameters of the CED HBT at peak  $f_T$ . (a) Width dependence for the frequency ratio  $f/f_T = 0.2, 0.5, 1, 1.5$ . (b) Frequency dependence for selected widths  $b_{E0}/\mu\text{m} = 0.2, 0.5, 1$ .

### 5.2.3 Large-signal transient operation

Due to its complexity, the investigation of fast large-signal switching has been a research topic for a long time. The early, strongly simplified, calculations of switching times in [12] were followed quickly by the introduction of the much more powerful charge-control theory [13]. The latter allowed not only to include charge storage elements and series resistances but also to calculate the time-dependent collector current. However,

closed-form solutions of the corresponding differential equation are only possible for bias independent elements and special cases of a forced base current as a function of time [14]. Charge-control theory was successfully employed until circuit simulators became widely used to allow a numerical solution of the transistor switching behavior for general time functions as well as nonlinear bias dependent elements.

The details of the switching behavior depend on the signal amplitude and shape over time (such as a step or a sinusoidal signal). For practical applications, signal amplitudes are (much) larger than  $V_T$ . As a consequence, the bias dependence of  $r_{SBI}$  and of the associated (area specific) total internal base charge,  $\bar{Q}_B$ , can no longer be neglected. Furthermore, particularly for bipolar transistors *fast* switching processes are of interest. If the corresponding signal is fast enough the bias dependence of  $r_{SBI}$  and  $\bar{Q}_B$ , caused by the large-signal change of the carrier densities in the base, results also in a spatial dependence of  $r_{SBI}$  and  $\bar{Q}_B$  as well as the current densities under the emitter. Therefore, lumped representations of the internal base region and its associated resistance, which are used in circuit simulators, become inadequate for fast switching. This was first indicated in [3][15][16] and later also demonstrated by 2D device simulation [17]. The essence of these investigations is illustrated schematically in Fig. 5.14 by means of the excess hole density in the neutral base along the boundary  $x_e$  to the BE SCR. Below, the mechanisms occurring in the internal transistor during high-speed switching are explained first qualitatively before possible modeling approaches and some quantitative results are being discussed.

During the turn-on process (Fig. 5.14a) the base current flows *into* the internal transistor region and supplies those holes that are necessary to charge the depletion capacitances and to compensate the injected minority carriers. Since the region close to the emitter edge at  $y = b_{E0}/2$  is charged first the potential there increases faster than in the center region ( $y = 0$ ). As a consequence, the BE and BC junction at the edge are more forward biased, which leads to a smaller sheet resistance and larger capacitance per area in this region compared to the center, causing  $r_{SBI}$  and  $\bar{Q}_B$  to become a function of  $y$ . In addition, the large charging current causes a significant voltage drop from the edge to the center (transient current crowding) and,

e.g., an electron current density that looks similar to the  $\Delta p$  curves in Fig. 5.14a.

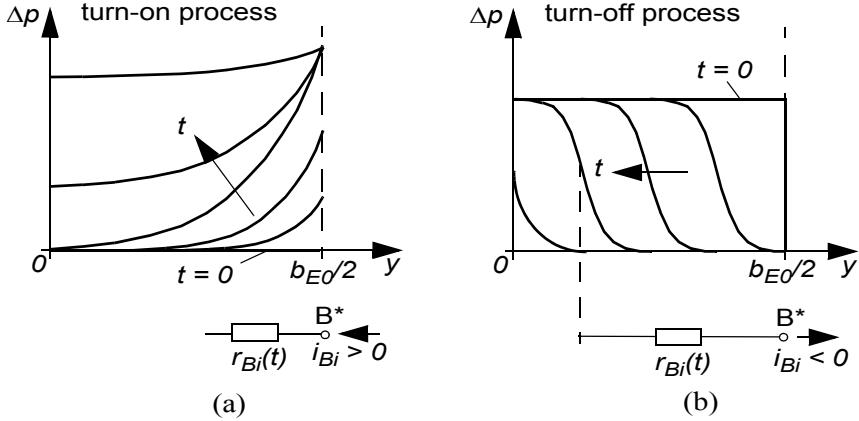


Fig. 5.14: Illustration of the excess hole density  $\Delta p(x_e)$  in the internal transistor during a high-speed switching process: (a) charging process of the internal transistor during turn-on; (b) discharge process during turn-off. The base contact is on the right, and a long stripe (2D case) is assumed.

The value of both  $r_{SBi}$  and  $\overline{Q}_B$  is determined by the evolution over time of the carrier density slope moving into center region. As long as the potential at the base node  $B^*$  increases sufficiently fast, the edge region remains significantly higher biased so that most of the holes carried by  $i_{Bi}(t)$  do not travel much further towards the center. Therefore, the internal base resistance  $r_{Bi}(t)$  remains smaller than the corresponding DC value, which is defined by an average internal base potential  $V_B$  and assuming a homogeneous distribution of  $\Delta p$  (and thus  $r_{SBi}$ ) over the emitter width. Once the signal change is slow enough the distribution of  $\Delta p$  and the value of  $r_{Bi}$  approach their respective DC values.

During the turn-off process (Fig. 5.14b) the base current flows *out of* the internal transistor region and carries those holes that both discharge the depletion regions and are no longer required to compensate the decreasing minority charge in the base. At the beginning, this discharge process occurs mostly at the emitter edge. Hence, the hole path is again very short, and the dynamic resistance  $r_{Bi}(t)$  is much smaller than the corresponding DC value. Over time though the edge region becomes more and more de-



pleted (since the bias across the pn junction has dropped to the turn-off value), so that increasingly charge in the regions towards the center is removed. This lead to a (negative) carrier density “wave” moving into the internal transistor. Now, not only the hole path becomes longer but the holes also have to travel through a depleted and, thus, high-ohmic region. As a result,  $r_{Bi}(t)$  increases rapidly and at some point in time exceeds the corresponding DC value. This can even be the case for most of the turn-off process if the input signal changes very rapidly. Again, from the above discussion it becomes obvious that  $r_{SBi}$  and  $\bar{Q}_B$  are strong functions of  $y$  during switching.

For an analytical treatment again the transistor chain in Fig. 5.4a is considered with the relevant input circuit elements shown in Fig. 5.15a. In this figure, the junction currents (diodes) have been neglected since for fast switching the capacitive current usually strongly dominates. However, for the sake of completeness the junction currents are included in the expressions below.

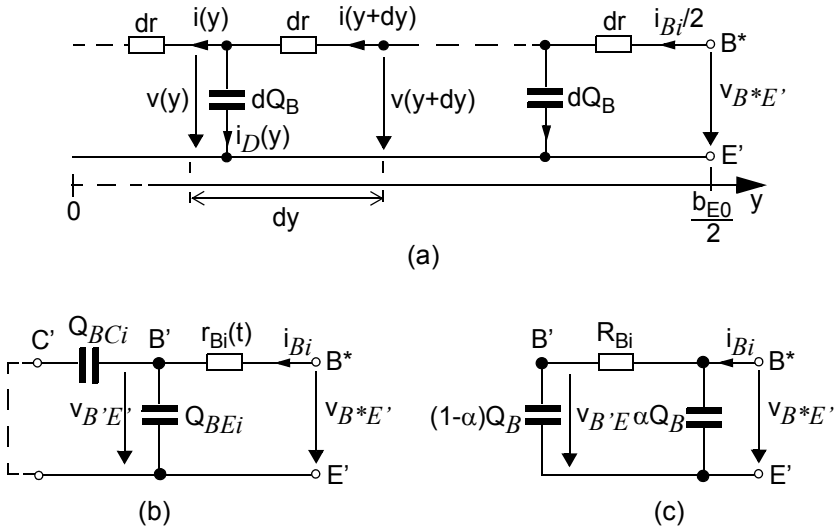


Fig. 5.15: Large-signal dynamic emitter current crowding during fast switching for a long (2D) transistor with two base contacts: (a) distributed representation, (b) lumped representation with transient  $r_{Bi}$  element, and (c) lumped representation with internal charge partitioned across the DC  $R_{Bi}$ . Assuming a CE-short,  $dQ_B(v) = \bar{Q}_B dy l_{E0}$  with  $\bar{Q}_B = \bar{Q}_{BEi} + \bar{Q}_{BCi}$ .

For the following considerations a dynamic CE short is assumed. The two differential equations that govern the lateral voltage drop and current continuity are

$$\frac{dv}{dy} = \frac{r_{SBi}(y)}{l_{E0}} i, \quad (5.59)$$

$$\frac{di}{dy} = l_{E0} \left[ J_{SBi} \exp\left(\frac{v(y)}{V_T}\right) + \frac{\partial \bar{Q}_B(y)}{\partial t} \right]. \quad (5.60)$$

Combining above equations and taking the spatial dependence of  $r_{SBi}$  into account yields

$$\frac{d^2 v}{dy^2} = r_{SBi} \left[ J_{SBi} \exp\left(\frac{v}{V_T}\right) + \frac{\partial \bar{Q}_B}{\partial t} \right] + \frac{i}{l_{E0}} \frac{dr_{SBi}}{dy}. \quad (5.61)$$

Assuming quasi-static conditions for the charging process in *vertical* direction, the time derivative of the charge can be written as

$$\frac{\partial \bar{Q}_B}{\partial t} = \frac{\partial \bar{Q}_B}{\partial v} \frac{\partial v}{\partial t} = \bar{C}_B(v) \frac{\partial v}{\partial t}. \quad (5.62)$$

Inserting this and the current from (5.59) into (5.61) leads to a nonlinear second-order partial differential equation for the voltage

$$\frac{d^2 v}{dy^2} - \left( \frac{1}{r_{SBi}} \frac{dr_{SBi}}{dy} \right) \frac{dv}{dy} - r_{SBi} J_{SBi} \exp\left(\frac{v}{V_T}\right) = r_{SBi} \bar{C}_B \frac{\partial v}{\partial t}. \quad (5.63)$$

Note, that the diode current term can be dropped for sufficiently fast switching.

In contrast to the DC and AC case, a solution of (5.63) for the complete switching process requires the knowledge of the time dependent input signal. In addition, for high-speed switching both  $r_{SBi}$  and  $\bar{Q}_B$  are a nonlinear function of  $v(y)$ . Therefore, a generic solution does not exist, and the analytical calculation of an effective base resistance is impossible. Only for very slow switching (i.e. spatially constant  $r_{SBi}$  and  $\bar{Q}_B$ ), which corresponds to DC conditions, and for extremely fast switching (infinite density slope) a solution exists [16]. In the latter case, however, still idealized assumptions apply, resulting in a 2-transistor model [18]. Usually though,

such a model cannot be realized in simulators as subcircuit, unless the bias dependence of  $r_{Bi}$  of the outer transistor element is negligible or the internal base node is externally accessible.

In an attempt to simplify (5.63) for compact modeling purposes, one notices that a decrease (increase) in  $\bar{Q}_B$  leads to an increase (decrease) of  $r_{SBi}$  at any  $y$ , hence compensating each other at least partially. Furthermore, during transient simulation (5.63) applies to any point in time. Thus, rather to try a solution for the whole switching process only the small time interval  $\delta t$  between two discrete points needs to be considered with the derivative  $\delta v/\delta t$  at the edge as an additional boundary condition for each iteration. However, the spatial derivative of  $r_{SBi}$  cannot be simplified for general signals, so that the differential equation remains nonlinear. It may be solved numerically at each time step, but at the expense of partially significant increase in computational effort.

For the compact representation in Fig. 5.15b, the definition of a lumped value for

$$r_{Bi}(t) = \frac{v_{B^*E'} - v_{B'E'}}{i_{Bi}} \quad (5.64)$$

during fast transient operation is more difficult since the diode current used to define  $v_{B'E'}$  in the DC definition is negligible and, hence, of no interest. Nevertheless,  $r_{Bi}(t)$  still defines an internal voltage  $v_{B'E'}$  that controls a lumped transfer current  $i_T(t)$  and hole charge  $Q_p(t)$  both of which are required to have the same time dependence as the corresponding variables of the distributed internal structure. Unfortunately, during switching  $i_T$  cannot be measured accurately via the terminal current  $i_C$  since the latter contains additional components. The charge at each time point can be calculated in device simulation and could be related to a  $V_{B'E'}$  value through its quasi-static relationship. However, for fast switching both  $Q_p$  and  $i_T$  contain also vertical NQS effects making any quasi-static operation based relation  $Q_p(V_{B'E'})$  or  $I_T(V_{B'E'})$  inaccurate. As a consequence, it appears to be impossible at this point to calculate the time dependent internal base resistance.

In the lumped representation shown in Fig. 5.15c the problem of having to find a special lumped formulation for  $r_{Bi}(t)$ , which may be incompatible with the result from HF AC analysis, is circumvented by partitioning the

charge of the internal transistor across the DC base resistance. The latter is still time dependent though via its regular bias dependence. Sometimes, it is suggested to integrate  $r_{Bi}(t)$  over the entire switching process to calculate an average value. While this yields the correct switching time it does not give the correct time dependence.

In order to provide a feel for realistic conditions during high-speed switching, device simulations of a turn-on and turn-off process were performed for the internal transistor structure shown in Fig. 5.16. The base contact is modelled the same way as for the 1D case. A smooth input signal  $v_{B^*E'}(t) = v_G(t)$  as shown in Fig. 5.17a,b was used to avoid artificial discontinuities.

Figures 5.17a,b contain the response of the collector and base current vs. time. It is evident that the switching base current is much larger than its DC value  $i_B(t=0)$  during the entire transient process. At the beginning of the transient process the collector current is slightly negative for turn-on and also slightly exceeds its DC value for turn-off due to the charging of the BC junction capacitance. Further delays of  $i_C(t)$  w.r.t the input voltage are caused by transient current crowding as well as by vertical non-quasi-static effects. The negative base current during turn-off results from discharging the transistor. It also indicates a voltage drop from  $B^*$  to  $B^*$ .

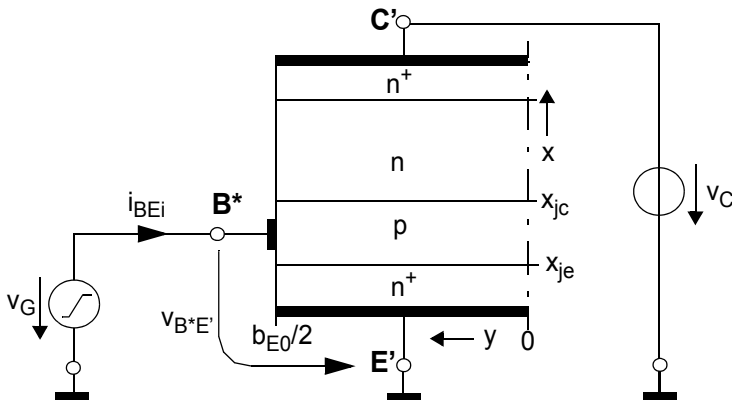


Fig. 5.16: Internal transistor structure and associated terminals used for subsequent theory discussion and device simulation. Only half of the structure is shown due to symmetry reasons.

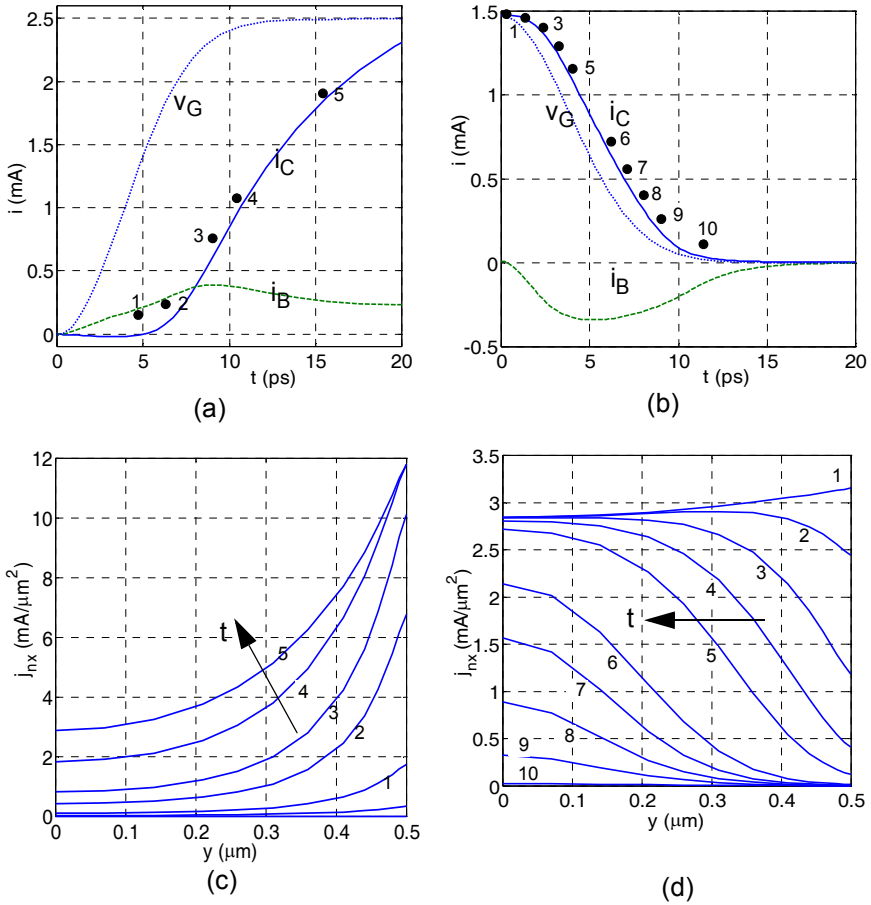


Fig. 5.17: Device simulation results of high-speed large-signal transient operation in the internal transistor. Time dependence of collector current (solid) and base current (dotted) for (a) turn-on process from  $V_{B^*E^*} = 0.7$  to  $0.95\text{V}$ , and (b) turn-off process from  $V_{B^*E^*} = 0.9$  to  $0.7\text{V}$ . For reference, the input signal  $v_G$  is shown in arbitrary units (dashed). Vertical electron current density at the BE junction vs. lateral location for selected time points: (c) turn-on process at  $t/\text{ps} = 4.8, 6.2, 8.6, 10.7, 15.7, 21.6$ , and (d) turn-off process at  $t/\text{ps} = 0, 1.1, 2.1, 3.1, 4.1, 6.1, 7.1, 8.1, 9.1, 11.1$ .

Transient current crowding is clearly visible in Fig. 5.17(c,d) from the laterally inhomogeneous distribution of the vertical electron current density (at the BE junction). The selected time points are marked in Fig. 5.17(a,b) in order to establish a correspondence to the terminal cur-

rent. Note that the voltage at the end of the turn-on process differs from the initial value of the turn-off process and, thus, also the voltage swing of the input signal in order to have a similar slope of  $i_C(t)$ . The final bias point of the turn-on process is located fairly far in the high-current region while the turn-off process starts just slightly beyond peak  $f_T$ .

The evolution of the electron current distribution is similar to that of the (excess) hole density, which was qualitatively described earlier. In both cases a very strong lateral dependence is visible during the transient input signal. The consequence of this strong spatial dependence is a significant electron current in lateral direction during turn-on that supports the equilibrium reached at the end of the switching process. This lateral current cannot be included in the transistor chain of Fig. 5.15 so that even such a sophisticated network model is only an approximation. The lateral electron current is negligible though during turn-off. For the example shown above a wide emitter was used. Similar results are obtained for narrow emitters (e.g. [19,20]), except that transient current crowding is less severe.

### 5.3 Emitter junction perimeter effects

In most processes, the emitter doping profile extends laterally beyond the emitter window edge during high temperature processing, leading to the emitter perimeter junction as sketched in Fig. 5.18. The result is a transistor structure at the perimeter the base width of which increases towards the external base contact. This is indicated in Fig. 5.18 by the corresponding flow lines of the injected transfer current. A physics-based lumped representation of the entire 2D structure consists of an internal transistor  $T_i$  carrying the current portion  $i_{Ti}$  and a perimeter transistor  $T_p$  carrying the current portion  $i_{Tp}$ . Also, the base current can be partitioned into an internal component  $i_{Bi}$  flowing into  $T_i$  and a perimeter component  $i_{Bp}$  flowing into  $T_p$ .

Furthermore, the injected currents carry an associated mobile charge component that is different for  $T_i$  and  $T_p$ . Moreover, the emitter perimeter junction itself leads to a depletion capacitance  $C_{jEp}$  (with its per area value varying along the junction), which differs from the capacitance  $C_{jEi}$  of the

internal junction. Finally, some portion of the external BC junction region is associated with the perimeter transistor.

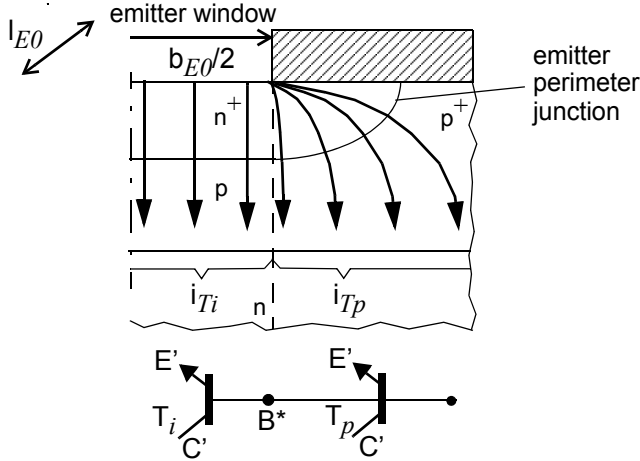


Fig. 5.18: Schematic cross-section of the internal and perimeter region. The arrows indicate the injected transfer current flow in the internal transistor ( $i_{Ti}$ ) and in the perimeter transistor ( $i_{Tp}$ ). At the bottom the resulting physics-based equivalent circuit is shown with a separate transistor for each region.

### 5.3.1 Standard scaling of currents and charges

The internal transfer current in Fig. 5.18 can be expressed as  $I_{Ti} = I_{T,A} A_{E0}$  with  $I_{T,A}$  as the internal current density (in, e.g.  $\text{mA}/\mu\text{m}^2$ ) and  $A_{E0} = b_{E0} l_{E0}$  as emitter window area. Since the perimeter junction area, given by the junction curvature, is unknown and cannot be measured, the perimeter current is more conveniently expressed as  $I_{Tp} = I_{T,P} P_{E0}$  with  $P_{E0} = 2(b_{E0} + l_{E0})$  as emitter window perimeter length and  $I_{T,P}$  as current per perimeter length (in, e.g.  $\text{mA}/\mu\text{m}$ ). In principle, instead of the emitter window any other dimensions may be chosen. However, the emitter window is generally the most suitable reference since it is a layout dimension and can also be measured using, e.g., electron microscopy.

The so-called *process-specific electrical* parameters  $I_{T,A}$  and  $I_{T,P}$  can be determined experimentally by measuring transistors with different emitter perimeter to area ratio (e.g. [21]), which corresponds to simply varying the emitter width in the 2D case. This simple separation approach into area and

perimeter component can be extended to other currents and charges associated with a junction. (Note that charges are generally measured through capacitances.) Therefore, be  $Y_e$  a total current, charge, conductance or capacitance, then according to Fig. 5.18 and assuming long stripes (i.e. the 2D case with  $l_0 \gg b_0$ )  $Y_e$  can be expressed by its specific parameters  $Y_{e,A}$  and  $Y_{e,P}$  in the form

$$Y_e = Y_{e,A}A_0 + Y_{e,P}P_0, \quad (5.65)$$

with  $A_0$  and  $P_0$ , respectively, as junction window area and perimeter length, respectively.

For compact modeling, the two components  $Y_{e,A}$  and  $Y_{e,P}$  have to be described as a function of bias (and temperature). While physics-based relations have already been derived for  $Y_{e,A}$  in the previous sections, this is more difficult for  $Y_{e,P}$  due to the non-planar geometry of the junction. In addition, at higher current densities the simple experimental separation of  $Y_{e,A}$  and  $Y_{e,P}$  mentioned above includes the impact of series resistances, current crowding and spreading as well as self-heating and high-current effects. For example, a voltage drop of just a few mV caused by a series resistance already leads to significant changes in a diode I(V) characteristic and the associated extracted model parameters. The unknown control voltages for  $Y_{e,A}$  and  $Y_{e,P}$  caused by the mentioned effects can only be corrected if the impact of these effects is accurately known in advance, which is generally not the case. Especially, the analytical description of the perimeter related transfer current and minority charge is very difficult at higher current densities.

As a consequence, the two-transistor equivalent circuit (EC) shown in Fig. 5.18 is generally not a suitable solution for a compact model. The complicated EC and associated computational effort is generally also not acceptable for circuit design purposes. A simple solution is to utilize (5.65) for defining an *effective electrical area* [22],

$$A = A_0 \left( 1 + \gamma \frac{P_0}{A_0} \right), \quad (5.66)$$

and to keep the bias dependent model equation of the 1D (i.e. specific area) component  $Y_{e,A}$  for describing the corresponding single EC element:



$$Y_e = Y_{e,A} A. \quad (5.67)$$

In other words, the per area (1D) model parameters are just multiplied with an effective area, which is given by the process-specific parameter

$$\gamma = Y_{e,P} / Y_{e,A} \quad (5.68)$$

and the layout dimensions.

This simple “standard-scaling” approach has been used very successfully for many generations of bipolar transistors over the past 25 years. However, for it to work properly the parameter  $\gamma$  must be independent on both dimensions and bias. In how far this is actually the case for *emitter perimeter junction* related effects is discussed in more detail below with the goal to derive solutions for a compact description of these effects.

### 5.3.2 Transfer current

The 1D GICCR is an accurate description of the vertical transfer current and includes many relevant bias dependent effects as well as structural details such as bandgap variations. However, real transistors exhibit to a more or less extent multi-dimensional effects. In a Si-based structure the most important effects impacting the assumptions for a 1D GICCR are emitter perimeter injection and collector current spreading. The influence of such multi-dimensional effects on the collector current formulation is discussed below, following the approach outlined first in [23]. For deriving the multi-dimensional GICCR it suffices here to focus first on the 2D case (i.e.  $l_{E0} \gg b_{E0}$ ), since a 3D structure does not contain any fundamental effects different from the respective 2D structure. For instance, possible corner effects can be described (and measured) using a 2D structure [24,25].

Figure 5.19 shows the relevant 2D cross-section. Due to symmetry reasons, only one half of the device is considered. The emitter contact is located within  $y = [0, b_{E0}/2]$  while the collector and base contact are located at  $x = x_C$ , and  $y = [b_{bc}, b_y]$ , respectively. The location  $x_C$  has already been defined in the 1D case, while  $b_y$  represents the total extension of the BC junction area between the symmetry line and, e.g., the shallow trench so that it includes the entire injected vertical electron current density.

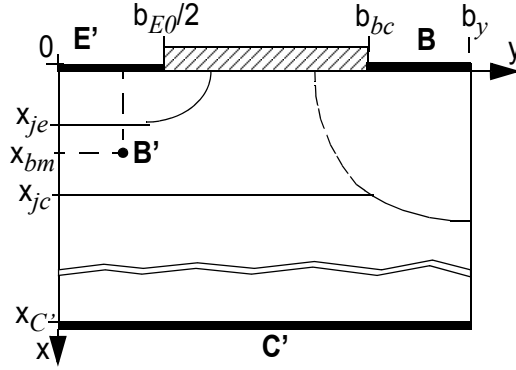


Fig. 5.19: Schematic cross-section of the 2D bulk structure considered in the derivation.

For deriving a multi-dimensional formulation of the GICCR, the following assumptions are being made:

- Negligible volume recombination, except in the emitter region.
- Quasi-static operation (equivalent to zero time derivative in the continuity equation).
- No loss of electron current at the base contact.
- Effects such as thermionic emission and tunneling across the junctions are neglected. They can be accounted for by separate terms.

As a consequence, it suffices to consider the vertical electron current density  $J_{nx}$  since eventually all electrons crossing the BE SCR will arrive at the contact  $C'$  (which is a line at  $x_{C'}$  in the 2D case). However, in contrast to the 1D case, now  $J_{nx}$  is in general a function of  $y$  due to current spreading and current crowding. The  $y$  dependence in turn causes  $J_{nx}$  also to vary with  $x$  via the continuity equation. This behavior is visualized in Fig. 5.20 by the electron current flow lines. Starting from the emitter center, each line contains an additional 10% of the vertical current flow (except for the last one which contains only 5%). The selected bias points correspond to low injection and high injection with a current density about 50% beyond that at peak  $f_T$ . These bias points will also be used in several subsequent figures.

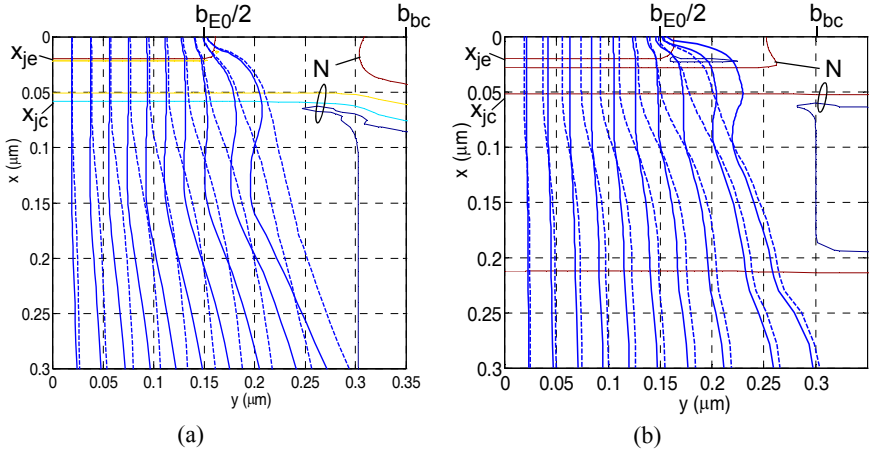


Fig. 5.20: Electron current flow lines for low-injection (solid lines) and for high injection (dashed lines): (a) BJT at  $I_C/A_{E0} = (0.75 \cdot 10^{-3}, 1.58) \text{ mA}/\mu\text{m}^2$  and  $V_{BE} = (0.7, 0.905) \text{ V}$ . (b) HBT at  $I_C/A_{E0} = (3 \cdot 10^{-3}, 7.63) \text{ mA}/\mu\text{m}^2$  and  $V_{BE} = (0.6, 0.925) \text{ V}$ .  $V_{B'C'} = 0 \text{ V}$ . Net doping contours have been inserted for reference.

For the BJT (Fig. 5.20a) the flow lines show very little bias dependence up to the BC junction. Just from counting the flow lines at the BE junction one can estimate the perimeter current to be about 20% of the total current. There are two distinct vertical regions of current spreading: the first one is in the (neutral) emitter and the second one is in the neutral collector. In the SCRs and neutral base, however, the larger drift field causes the carriers to flow perpendicular to the junctions. At low injection, both the high electric field in the BC SCR and, to a lesser extent, also the lateral collector doping difference of the SIC focus the flow lines into x-direction. At high injection the collapse of the electric field in the BC SCR can be clearly observed through the increased current spreading in the collector just beyond the BC junction.

For the HBT (Fig. 5.20b) the low current perimeter injection is larger than in the BJT due to the much lower doped base (cap) layer at the surface of the spacer region. However, as this cap layer gets swamped by carriers at high injection, conductivity modulation reduces the increase of current with  $V_{BE}$  compared to the injection across the internal BE junction. Hence, the perimeter current contribution decreases at high injection. In the col-

lector the collapse of the electric field at the BC junction is again clearly visible by the increased current spreading.

The derivation of the 2D GICCR starts with the extension of (4.33) by the reference value  $\mu_{nr}n_{ir}^2$  as well as by the desired variable  $I_T/A_{E0}$  and the term  $\exp(V_{B'E}/V_T)$ , which yields after rearrangement

$$I_T h(x, y)p(x, y) = -c_0 \frac{d[\exp(-\phi_n/V_T)]}{dx} \exp\left(\frac{V_{B'E}}{V_T}\right) \quad (5.69a)$$

with the constant

$$c_0 = qV_T A_{E0} \mu_{nr} n_{ir}^2. \quad (5.69b)$$

Furthermore, the normalized 2D weight function

$$h(x, y) = h_g(x, y)h_J(x, y)h_v(x, y) \quad (5.70a)$$

consists, like in the 1D case, of the components

$$h_g = \frac{\mu_{nr}n_{ir}^2}{\mu_n n_i^2}, \quad h_J = \frac{-J_{nx}}{I_T/A_{E0}}, \quad h_v = \exp\frac{V_{B'E} - \phi_p}{V_T}, \quad (5.70b)$$

which now depend also on  $y$  and will be discussed later in more detail.

Integration over the entire 2D transistor region (taking into account lateral symmetry) gives

$$I_T \int_0^{b_y} \bar{Q}_{ph}(y) dy = -qc_0 \exp(V_{B'E}/V_T) \int_0^{b_y} S_\phi(y) dy \quad (5.71a)$$

with the area specific (1D) weighted hole charge

$$\bar{Q}_{ph}(y) = q \int_0^{x_C} h p dx \quad (5.71b)$$

and the 1D integral related to the QFP,

$$S_\phi(y) = \int \frac{\exp\left(\frac{\phi_n(x_C, y)}{V_T}\right)}{\exp\left(\frac{\phi_n(0, y)}{V_T}\right)} d[\exp(-\phi_n/V_T)]. \quad (5.71c)$$

The integral expressions in  $x$  direction are the same as in the 1D case, while in  $y$  direction the interval extends laterally as far as a significant contribution of  $J_{nx}$  exists. Practically, the interval  $[-b_y, b_y]$  is given by width of the

BC junction within either the inner shallow trench edges or the current flow lines that include close to 100% of the electron current. The integrals in above relation are discussed below separately.

The r.h.s. integral is evaluated first in  $x$  direction (cf. (4.34)):

$$S_{\varphi}(y) = \exp\left(-\frac{\varphi_n(x_{C'}, y)}{V_T}\right) - \exp\left(-\frac{\varphi_n(0, y)}{V_T}\right). \quad (5.72)$$

At the lines  $x = 0$  and  $x = x_{C'}$ , the QFP  $\varphi_n$  assumes the values of the respective applied voltages within the contact regions. Choosing again, as in the 1D case,  $x = 0$  as potential reference gives  $\varphi_n(0, y) = 0$  within the emitter window  $0 \leq y \leq b_{E0}/2$  (cf. Fig. 5.19) as well as  $\varphi_n(x_{C'}, y) = V_{C'E'}$  along the entire epi-collector to buried layer transition line ( $0 \leq y \leq b_{bc}$ ). Between the emitter contact edge at  $y = b_{E0}/2$  and the adjacent base contact edge at  $b_{bc}$ , the potential  $\varphi_n(0, y)$  increases from 0 to  $V_{BE'}$ ; i.e. the potential at the point  $(0, b_{bc})$  equals that of the base terminal B. In contrast,  $V_{B'E'}$  and  $V_{C'E'}$  are the *internal* voltages, which are smaller than the terminal voltages due to the voltage drops across the remaining external series resistances of the 2D structure. Inserting the above boundary conditions into (5.72) yields after performing the integration

$$b_E = \int_0^{b_y} S_{\varphi}(y) dy = \exp\left(-\frac{V_{C'E'}}{V_T}\right) b_y - \left(\frac{b_{E0}}{2} + \frac{\Delta b_E}{2}\right) \quad (5.73)$$

with the definition of the lateral emitter window width extension

$$\frac{\Delta b_E}{2} = \int_{b_{E0}/2}^{b_{bc}} \exp\left(-\frac{\varphi_n(0, y)}{V_T}\right) dy + (b_y - b_{bc}) \exp\left(-\frac{V_{BE'}}{V_T}\right). \quad (5.74)$$

Outside the emitter window,  $\varphi_n(0, y)$  increases within the BE spacer region towards  $V_{BE'}$ . Hence, at any *reasonable forward bias* the value of  $\exp(-\varphi_n(0, y)/V_T)$  drops rapidly and the last term in (5.74) is negligible. The spatial dependence of the contribution to the integral in (5.74) is shown in Fig. 5.21 for three different bias points of the BJT and HBT. The medium bias point corresponds to peak  $f_T$ . Although the contribution to the integral continues into the spacer region, it is evident that the extension  $\Delta b_E$  of the emitter is generally significantly smaller than  $b_{E0}$ . The resulting dimension

$$b_E = b_{E0} + \Delta b_E \quad (5.75)$$

will be referred to as *effective electrical emitter width* that includes the effect of emitter perimeter injection. For the BJT (Fig. 5.21a) the extension  $\Delta b_E$  is hardly bias dependent. In contrast, the strong decrease of  $\Delta b_E$  in the HBT (Fig. 5.21b) is caused by the previously mentioned bias dependence of the perimeter injection into the base cap layer.

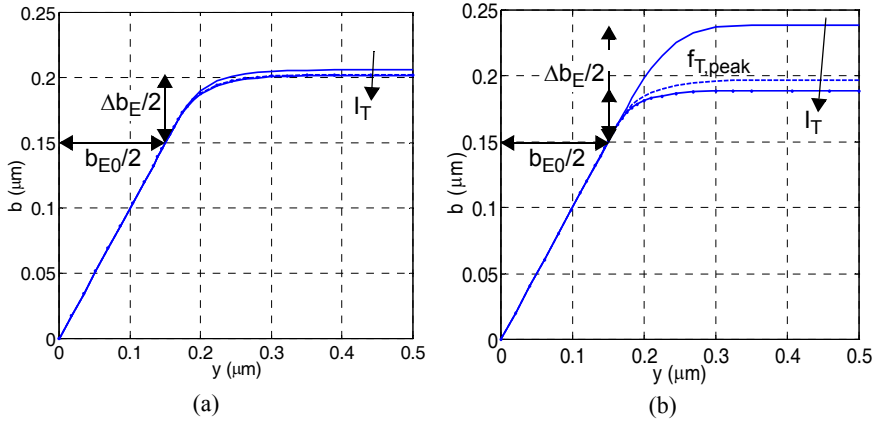


Fig. 5.21: Effective electrical emitter width (5.73) from 2D device simulation at different bias points: (a) BJT at  $I_C/A_{E0} = (0.75 \cdot 10^{-3}, 0.96, 1.58) \text{ mA}/\mu\text{m}^2$ ; (b) CED HBT at  $I_C/A_{E0} = (3 \cdot 10^{-3}, 3.9, 7.63) \text{ mA}/\mu\text{m}^2$ .  $V_{BE'} = 0\text{V}$ .

As the next step the integral on the l.h.s. of (5.71a),

$$Q_{ph} = \int_0^{b_y} \bar{Q}_{ph}(y) dy = q \int_0^{b_y} \int_0^{x_{C'}} h_g h_J h_v p dx dy, \quad (5.76)$$

shall be investigated in more detail, starting with the different weight functions. For  $h_g$  similar considerations as for the 1D case apply for the region under the emitter.

The function  $h_v$  behaves under the emitter similarly to the 1D case as long as DC emitter current crowding can be neglected. In other words,  $h_v$  is close to 1 at low current densities assuming that  $V_{BE'}$  is defined by  $\phi_p$  at an appropriate location  $(x_B', y_B')$  in the internal base region. At high current densities, the voltage drop across the internal base resistance causes

$\phi_p$  under the emitter to drop compared to the external base. Hence  $h_v$  becomes larger than 1 at locations  $y < y_B$  and smaller than 1 for  $y > y_B$ .

In contrast to the 1D case, the function  $h_J$  is now spatially dependent on  $x$  and  $y$  due to collector current spreading and, at high current densities, also emitter current crowding (in the base). According to Fig. 5.20,  $J_{nx}$  and hence  $h_J$  fade rapidly outside the internal transistor (especially in the BE region) so that the hole density in the external base is weighted with very small values. In the collector the current spreading is larger but the hole density is very small there, especially in the external collector region, so that the latter contributes very little to the integral.

Figure 5.22 provides a quantitative overview on the lateral spatial dependence of the weight functions at the fixed location  $x_{bm}$  in the neutral base along with the normalized zero-bias hole charge density. Since under the emitter the possible  $y$  dependence of  $J_{nx}$  is mainly caused by emitter current crowding, which in turn is proportional to  $\exp(\phi_p/V_T)$ , multiplication of  $h_J$  with  $h_v$  should lead to a laterally constant value of the combined weight function

$$h_{Jv} = h_J h_v \quad (5.77)$$

under the emitter and is indeed observed in Fig. 5.22. The drop towards the external base is caused almost entirely by  $h_J$ . For the BJT (Fig. 5.22a), the bias dependence of  $h_{Jv}$  is very small (as already observed in the flow lines before), while for the HBT (Fig. 5.22b)  $h_{Jv}$  behaves more one-dimensional at higher injection. Note that  $h_{Jv} < 1$  under the emitter since its normalization factor contains  $A_{E0}$  instead of the effective emitter dimensions. As expected,  $h_g$  varies only little throughout the internal base and emitter perimeter region and then drops in the external base due to the high base doping there.

As Fig. 5.22 clearly visualizes,  $h_{Jv}$  ultimately determines the lateral integration interval in the 2D (and 3D) case. In other words, *a hole contribution comes only from those regions, where (vertical) transfer current flow is present*. Especially, hole contributions in the external base, where the hole density can become very large due to the high doping concentration, are suppressed by  $h_J$  and contribute very little to the overall integral.

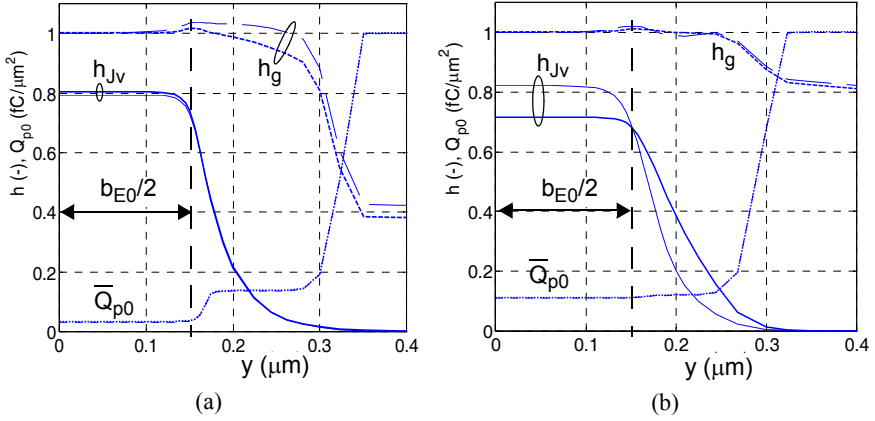


Fig. 5.22: Lateral ( $y$ ) dependence of the weight functions  $h_{JV}$  and  $h_g/h_g(x_{bm}, 0)$  along with the normalized hole charge density  $\bar{Q}_{p0}/\bar{Q}_{p0}(b_y)$  at a fixed value  $x = x_{bm}$ : (a) BJT for  $x_{bm} = 0.04 \mu\text{m}$  and at  $I_C/A_{E0} = (0.75 \cdot 10^{-3}, 1.58) \text{ mA}/\mu\text{m}^2$ ; (b) HBT for  $x_{bm} = 0.04 \mu\text{m}$  and at  $I_C/A_{E0} = (3 \cdot 10^{-3}, 7.63) \text{ mA}/\mu\text{m}^2$ .  $V_{BE} = 0\text{V}$ . Thick (thin) lines represent the low (high) current density.

As a different perspective, Fig. 5.23 shows the *depth* dependence of  $h_{JV}$  together with the change of the weighted hole density w.r.t. equilibrium,  $h_g \Delta p$ , at different lateral locations  $y$  and for the two previously selected bias points. At low injection,  $h_g \Delta p$  is concentrated to narrow regions around the junctions, while at high injection it significantly spreads towards both emitter and collector. According to Fig. 5.22 and 1D analysis, both  $h_g$  as well as  $h_g \Delta p$  depend significantly on  $x$ , but only weakly on  $y$ . In contrast, the shape of  $h_{JV}$  depends only weakly on  $x$ , especially under the emitter, despite the very different current densities. Therefore, in the subsequent analysis the weight functions are replaced by the approximations

$$h_{JV,y} = h_{JV}(x_{bm}, y), \quad h_{g,x} = h_g(x, 0). \quad (5.78)$$

The additional index indicates the spatial dependence. Note that these factors also are still a function of bias.



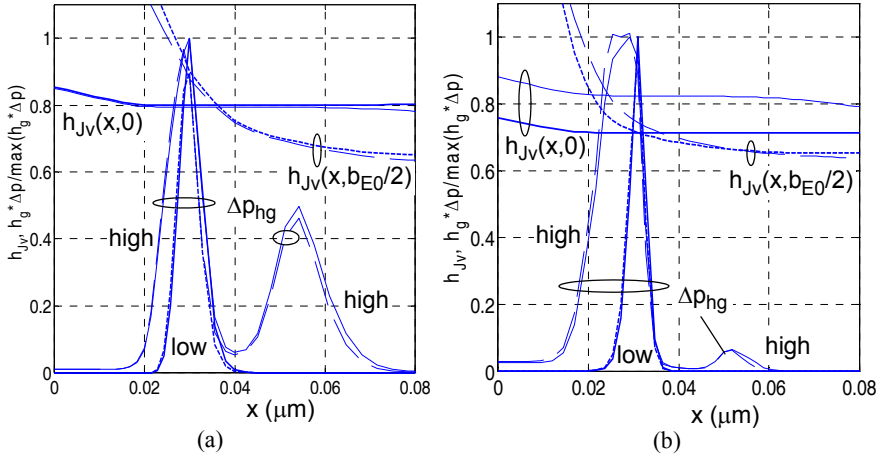


Fig. 5.23: Depth ( $x$ ) dependence of weight function  $h_{JV}$  and weighted hole density change  $\Delta p_{hg} = h_g \Delta p$  at fixed values of  $y = 0$  (solid lines) and  $y = b_{E0}/2$  (dashed lines): (a) BJT at  $I_C/A_{E0} = (0.75 \cdot 10^{-3}, 1.58) \text{ mA}/\mu\text{m}^2$  and (b) HBT at  $I_C/A_{E0} = (3 \cdot 10^{-3}, 7.63) \text{ mA}/\mu\text{m}^2$ .  $V_{B-C} = 0\text{V}$ . Thick (thin) lines represent low (high) injection.

As a consequence of the 2D integration, the definition and meaning of the hole charge, which is simple in the 1D case, needs to be revisited. For the evaluation of (5.76), as a first step, the hole density is split, as in the 1D case, into the equilibrium component  $p_0(x,y)$  and the bias dependent component  $\Delta p(x,y)$ , taken with respect to the equilibrium condition:

$$\int_0^{b_y} \bar{Q}_{ph}(y) dy = Q_{p0h} + \Delta Q_{ph}. \quad (5.79)$$

The first term on the r.h.s. leads to the weighted zero-bias hole charge of the 2D structure,

$$Q_{p0h} = 2l_{E0}q \int_0^{b_{bc}} \int_0^{x_{C'}} h p_0 dx dy, \quad (5.80)$$

which obviously now depends also on the lateral distribution of the weight function  $h(x,y)$ . The second expression on the r.h.s. of (5.79) corresponds to the *weighted change* of hole charge w.r.t. equilibrium,

$$\Delta Q_{ph} = 2l_{E0}q \int_0^{b_{bc}} \int_0^{x_{C'}} h \Delta p dx dy. \quad (5.81)$$

Inserting these charges and the effective electrical emitter width from (5.73) into (5.71a) yields for the transfer current

$$I_T = ql_{E0}b_{Ec0} \frac{\left[ \exp\left(\frac{V_{B'E'}}{V_T}\right) - 1 \right] - \left[ \exp\left(\frac{V_{B'C}}{V_T}\right) - 1 \right] \frac{2b_y}{b_E}}{Q_{p0h} + \Delta Q_{ph}}. \quad (5.82)$$

The “-1” in the brackets results from (5.74) and ensures the correct solution at low and zero bias. Evaluating (5.82) using internal solution variables from 2D device simulation yields excellent agreement with the terminal (collector) current from device simulation as shown in Fig. 5.24. This is expected since, except for neglecting the recombination, (5.82) is an exact solution for the quasi-static case and, hence, will serve as reference for further simplifications towards practical applications, particularly for deriving a geometry scalable formulation for  $I_T$ .

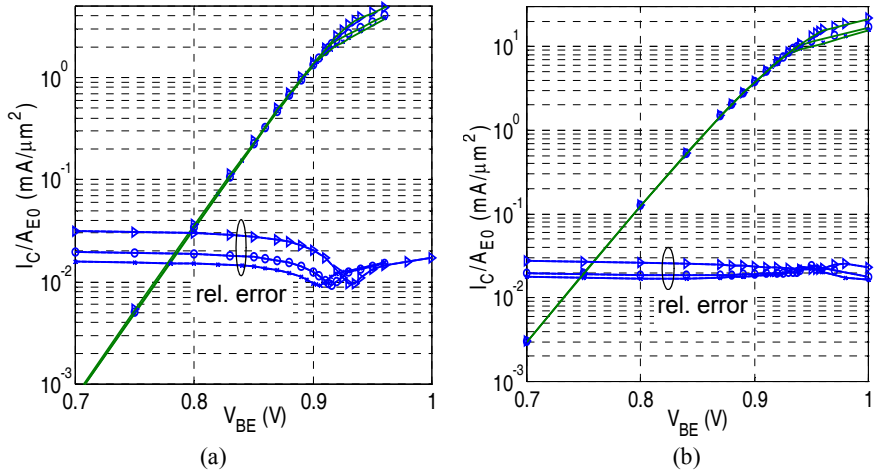


Fig. 5.24: Transfer current from (5.82) and from device simulation vs. BE terminal voltage at  $V_{BC} = (0.4, 0, -2.3)$  V for (a) BJT and (b) HBT. The relative error is below 0.4% and is caused mostly by the numerical integration on the discrete simulation grid.

Note, that according to the derivation of (5.82) the choice of  $V_{B'E'}$  is arbitrary. However, in the context of a compact model this voltage is defined by the base resistance and base terminal voltage. Thus, the comparison in Fig. 5.24 is performed for the voltage  $V_{B'E'} = \varphi_p(x_{Bm}, y_{B'})$  with  $y_{B'} = (2/3)b_{E0}/2$  which corresponds to the center of gravity of the lateral base cur-

rent distribution at negligible emitter current crowding. Relation (5.82) also shows that, as a consequence of the non-symmetrical structure, the operation of a bipolar transistor is not symmetric with respect to  $V_{B'E'}$  and  $V_{B'C'}$ . For a compact model, the weighted charges need to be expressed by measurable variables. This will be done below separately first for  $Q_{p0h}$  and then for  $\Delta Q_{ph}$ .

With the assumption (5.78) one can write (5.80) as

$$Q_{p0h} \cong 2l_{E0} \int_0^{b_y} h_{Jv,y} \left( q \int_0^{x_{C'}} h_{g,x} p_0 dx \right) dy. \quad (5.83)$$

According to Fig. 5.22 the inner integral can be split into an internal and an external contribution,

$$Q_{p0h} \cong 2l_{E0} \left[ \int_0^{\frac{b_{E0}}{2}} h_{Jv,y} \bar{Q}_{p0h} dy + \int_{\frac{b_{E0}}{2}}^{b_y} h_{Jv,y} q \int_0^{x_{C'}} h_{g,x} p_0 dx dy \right], \quad (5.84)$$

with the weighted 1D zero-bias hole charge  $\bar{Q}_{p0h}$  from (4.54). For HBTs (cf. Fig. 5.22b) the spacer charge density equals  $\bar{Q}_{p0h}$  so that

$$Q_{p0h} \cong 2l_{E0} \bar{Q}_{p0h} \int_0^{b_y} h_{Jv,y} dy. \quad (5.85)$$

With the definition (5.70b) for  $h_j$  it can be easily shown that the remaining integral in (5.85) equals  $b_{E0}/2$  as long as emitter current crowding is negligible. Even if this is not the case, the integral value still differs only slightly from  $b_{E0}/2$ . As a result, one obtains

$$Q_{p0h} \cong l_{E0} b_{E0} \bar{Q}_{p0h}. \quad (5.86)$$

For BJTs (cf. Fig. 5.22a) the inner integral of the second term in the brackets of (5.84) is typically larger than  $\bar{Q}_{p0h}$  since the doping peak is closer to the surface. Assuming simply  $r\bar{Q}_{p0h}$  with  $r > 1$  within the spacer region, one obtains

$$Q_{p0h} \cong l_{E0} b_{E0} \bar{Q}_{p0h} \left[ 1 + (r-1) \frac{I_{Tp}}{I_T} \right] \quad (5.87)$$

with  $I_{Tp}$  as the injected perimeter current, which clearly defines the amount of contribution from the hole charge under the BE spacer and in

the external base. Inserting the 1D zero-bias hole charge density  $\bar{Q}_{p0}$ . Therefore, one can generally write

$$Q_{p0h} \cong \bar{h}_{02} A_{E0} \bar{Q}_{p0}. \quad (5.88)$$

Here,  $\bar{h}_{02}$  includes both the lateral *and* vertical average weight factor.

Figure 5.25 shows the current dependence of  $\bar{h}_{02}$  for different  $V_{BC}$  values. For BJTs the distribution of  $h_{jv}$  in the base region, where  $p_0$  contributes the most, is almost bias independent. Thus the bias dependence of  $Q_{p0h}$  is mainly given by 1D theory which is confirmed by comparison with Fig. 4.45. In HBTs the distribution of  $h_{jv}$  in the base region can change with current, causing  $Q_{p0h}$  to start decreasing already at medium injection.

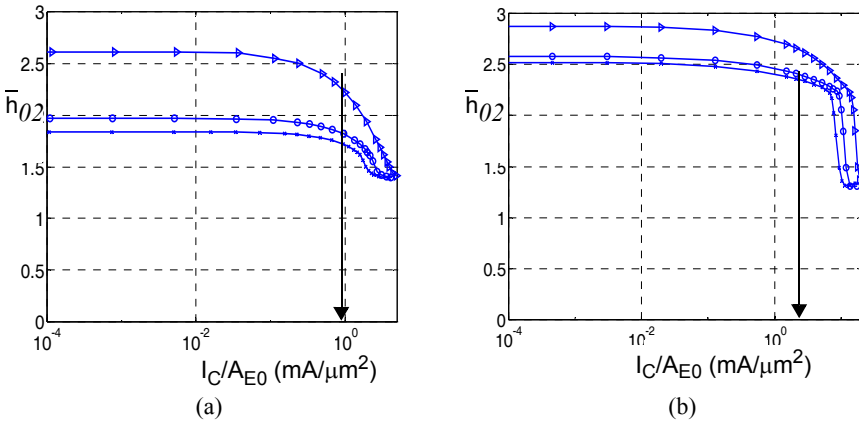


Fig. 5.25: Weighted zero-bias charge  $Q_{p0h}$ , normalized to the actual charge  $A_{E0}\bar{Q}_{p0}$ , vs. collector current density for (a) BJT and (b) HBT.  $V_{BC}/V = 0.4$  (x), 0 (O), -2.3 (Δ). The arrows indicate peak  $f_T$  at  $V_{BC}/V = 0V$ .

The next step is to express  $\Delta Q_{ph}$  by a measurable variable. In order to keep the 1D formulation, an attractive option is to establish a link between  $\Delta Q_{ph}$  and the measurable *actual* hole charge (i.e. its change) during transient and small-signal operation,

$$\Delta Q_p = 2l_{E0}q \int_0^{b_y} \int_0^{x_C} \Delta p \, dx \, dy. \quad (5.89)$$

As was shown for the 1D case in section. 4.3.1.3, the individual charges need to be weighted with factors associated with the vertical charge region since both high-doping effects and material composition can cause the in-

intrinsic carrier density to vary significantly over the depth of the transistor. Following the 1D approach, the 2D charge  $\Delta Q_p$  is partitioned into its depletion and minority charge components,

$$\Delta Q_p = Q_{jE} + Q_{jC} + Q_m . \quad (5.90)$$

However, now  $Q_{jE}$  and  $Q_{jC}$  are the *total* depletion charges associated with the internal and external portion of the junctions, respectively, and  $Q_m$  is the *total* minority charge.

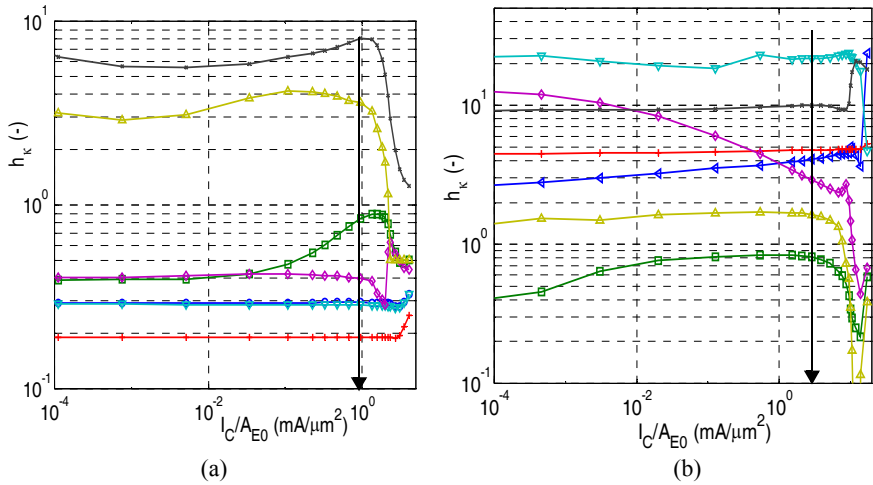


Fig. 5.26: Current dependence of the normalized weight factors defined in (5.92) for the various charge components, obtained from device simulation for  $V_{B'C'} = 0\text{V}$ : (a) BJT and (b) CED HBT. Legend:  $\bar{h}_{jE}(\circ)$ ,  $\bar{h}_{jC}(\square)$ ,  $\bar{h}_{fE}(+)$ ,  $\bar{h}_{fBE}(\nabla)$ ,  $\bar{h}_{fB}(\diamond)$ ,  $\bar{h}_{fBC}(\Delta)$ ,  $\bar{h}_{fC}(\times)$ . The arrow indicates peak  $f_T$ .

Applying the Regional Approach to the 2D structure leads, as in the 1D case, to charge components and their weighted counterparts in the neutral emitter, the BE SCR, the neutral base, the BC SCR and the neutral collector. Compared to the 1D case these components now possess a lateral dependence and are weighted mainly with  $h_{J_V}$ . The exact 2D weight factors are defined for each charge component as

$$\bar{h}_{\kappa 2} = \Delta Q_{\kappa h} / \Delta Q_\kappa \quad (5.91)$$

with  $\Delta Q_\kappa$  and  $\Delta Q_{\kappa h}$  as the total actual and weighted component in the respective spatial region  $\kappa$ . Normalizing these, as in the 1D case, to the low current value  $\bar{h}_{02,l}(V_{B'C'} = 0)$  of (5.88) gives

$$\bar{h}_\kappa = \bar{h}_{\kappa 2} / \bar{h}_{02,l} \quad (5.92)$$

with  $\kappa = (0, jE, jC, fE, fBE, fB, fBC, fC)$ . Figure 5.26 shows the bias dependence of these weight factors obtained from evaluating the results of 2D device simulation. (Note, that  $\bar{h}_{02}$  was already displayed in Fig. 5.25). For the BJT most factors, except  $\bar{h}_{jC}$  and  $\bar{h}_0$  can be assumed as constant up to current densities beyond peak  $f_T$ . For the HBT, additionally  $\bar{h}_{jE}$  and  $\bar{h}_{fB}$  are current dependent due to the larger variation of  $h$  in vertical direction.

Expressing the weighted charge (5.79) by its components

$$Q_{p0h} + \Delta Q_{ph} = \bar{h}_{02} A_{E0} \bar{Q}_{p0} + \sum_{\kappa} \bar{h}_{\kappa 2} A_{E0} \Delta \bar{Q}_{\kappa} \quad (5.93)$$

yields after inserting into (5.82) the general formulation

$$I_T = I_{TS} \frac{\exp\left(\frac{V_{B'E'}}{V_T}\right) - \exp\left(\frac{V_{B'C'}}{V_T}\right) \frac{2b_y}{b_E}}{\bar{h}_0 + \sum_{\kappa} \bar{h}_{\kappa} \frac{\bar{Q}_{\kappa}}{\bar{Q}_{p0}}} \quad (5.94a)$$

with  $\bar{h}_0 = \bar{h}_{02} / \bar{h}_{02,l}$ . Since the denominator has been normalized to the constant zero-bias charge factor  $\bar{h}_{02,l} A_{E0} \bar{Q}_{p0}$  the saturation current

$$I_{TS} = \frac{q^2 V_T \mu_{nr} n_{ir}^2}{\bar{h}_{02,l} \bar{Q}_{p0}} l_{E0} b_E = J_{TS} l_{E0} b_E. \quad (5.94b)$$

does not depend on bias as long as  $b_E$  is constant.  $I_T$  scales directly with the effective electrical emitter width (or area in the 3D case) contained in the saturation current. In addition, the charge weight factors can include a significant geometry dependence. The sum in (5.94a) contains the bias dependent actual charge densities, which are modelled according to 1D theory and weighted according to their different regions and 2D/3D effects.

Figure 5.27 provides a feel for the impact of the different components and their weight factors on  $I_T$ . Each curve represents the relative error of  $i_T$  according to (5.94a) if one weight factor is kept constant at its low-current value. The most significant influence for both transistor types and up

to current densities significantly beyond that of peak  $f_T$  stems from the bias dependence of  $b_E$  and  $\bar{h}_0$ . For the BJT also  $\bar{h}_{jC}$  is important, which represents mostly the mobile (majority) charge in the BC SCR since the curves in Fig. 5.27 are valid for  $V_{B'C'} = 0$ , i.e. zero fixed depletion charge. For the HBT, however,  $\bar{h}_{jE}$  and  $\bar{h}_{fB}$  are relevant at low and medium injection. At very high injection, assuming constant factors  $\bar{h}_{fBC}$  and  $\bar{h}_{fB}$  for the minority (electron) charge in the base portion of the BC SCR and the neutral base, respectively, as well as in particular a constant factor  $\bar{h}_{fC}$  for the charge injected into the collector cause larger errors. It is questionable though whether accurate modeling of these factors is required for practical applications.

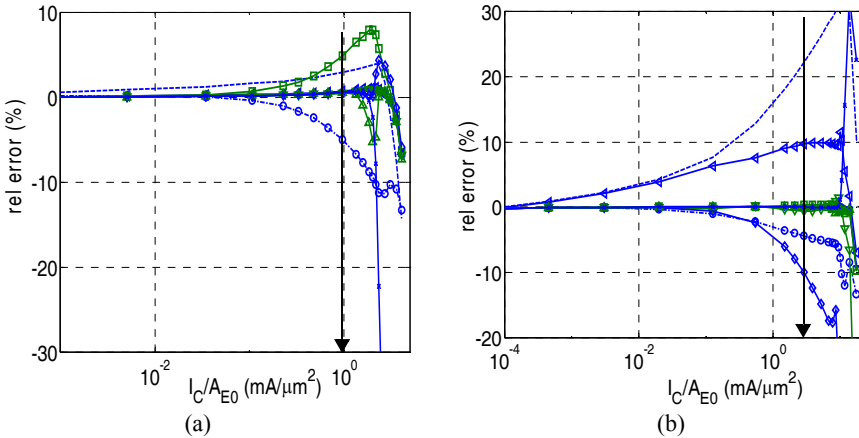


Fig. 5.27: Relative error of transfer current from the (5.94a) with the weight factors and charge components taken directly from device simulation using the regional approach: (a) BJT, (b) HBT.  $V_{B'C'} = 0V$ . For each curve only *one* weight factor has been set constant:  $\bar{h}_0(o)$ ,  $\bar{h}_{jE}(<)$ ,  $\bar{h}_{jC}(\square)$ ,  $\bar{h}_{fE}(+)$ ,  $\bar{h}_{fBE}(\nabla)$ ,  $\bar{h}_{fB}(\diamond)$ ,  $\bar{h}_{fBC}(\Delta)$ ,  $\bar{h}_{fC}(x)$ . The dashed line represents the influence of the effective emitter width  $b_E$ , and the arrow indicates peak  $f_T$ .

Detailed analytical descriptions for the weighted charges depend on the specific compact model and are beyond the scope of this chapter. However, the impact of 2D effects on a few important components and a generally possible analytical approach are considered below.

The BE depletion charge *density* and its weighted counterpart,

$$\Delta \bar{Q}_{jEh} \cong \int_{x_{mE}}^{x_{Bn}} h_g \Delta \rho dx, \quad (5.95)$$

generally vary laterally along the junction. In addition, the integration limits become  $y$  dependent at the emitter periphery. Taking the weighted internal charge density  $\Delta \bar{Q}_{jEih} = \Delta \bar{Q}_{jEh}(y=0)$  as reference and making the reasonable assumption that  $\Delta \bar{Q}_{jEih}$  is constant over the emitter window width gives with (5.78)

$$\Delta Q_{jEh} \cong 2l_{E0} \Delta \bar{Q}_{jEih} \left[ \int_0^{b_{E0}/2} h_{Jv,y} dy + \int_{b_{E0}/2}^{b_y} h_{Jv,y} \frac{\Delta \bar{Q}_{jEh}}{\Delta \bar{Q}_{jEih}} dy \right]. \quad (5.96)$$

$\Delta \bar{Q}_{jEih}$  is related to the internal BE space charge density  $Q_{jEi}/A_{E0}$  according to the 1D theory (4.69) where the weight factor  $\bar{h}_{BE}$  is given by  $h_g$  only. The value of the second integrand at the emitter periphery depends on the process. For BJTs, usually the base doping concentration increases towards the surface leading to  $\Delta \bar{Q}_{jEh}/\Delta \bar{Q}_{jEih} > 1$ . In HBTs, a retrograde base profile exists in the spacer region of the perimeter junction, leading to  $\Delta \bar{Q}_{jEh}/\Delta \bar{Q}_{jEih} < 1$ . Thus the charge density ratio leads to a contribution from the perimeter, and the term in the brackets is just equivalent to an effective cross-section that is larger than  $b_{E0}$ . After inserting (4.69) this leads to

$$\Delta Q_{jEh} \cong \bar{h}_{jE2} A_{E0} \bar{Q}_{jEi}, \quad (5.97)$$

in which the weight factor

$$\bar{h}_{jE2} = \bar{h}_{BE} \frac{2}{b_{E0}} \left[ \int_0^{b_{E0}/2} h_{Jv,y} dy + \int_{b_{E0}/2}^{b_y} h_{Jv,y} \frac{\Delta \bar{Q}_{jEh}}{\Delta \bar{Q}_{jEih}} dy \right] \quad (5.98)$$

now has absorbed both the 1D and 2D component, and  $\bar{Q}_{jEi}$  is the measurable 1D BE depletion charge density. The assumption of laterally independent  $\Delta \bar{Q}_{jEih}$  and  $\bar{Q}_{jEi}$  is justified up to medium current densities, i.e. as long as this charge plays a significant role in  $\Delta Q_p$ .

The BC depletion charge can be treated along the same lines as above leading to an average weight factor

$$\bar{h}_{jC2} = \bar{h}_{BC} \frac{2}{b_{E0}} \left[ \int_0^{b_{E0}/2} h_{Jv,y} dy + \int_{b_{E0}/2}^{b_y} h_{Jv,y} \frac{\Delta \bar{Q}_{jCh}}{\Delta \bar{Q}_{jCih}} dy \right]. \quad (5.99)$$



The effective cross-section given by the terms in brackets equals  $b_{E0}/2$  if the SIC is implanted through the base poly window so that  $\Delta\bar{Q}_{jCh} = \Delta\bar{Q}_{jCih}$  under the spacer. However, if the SIC is implanted just through the emitter window  $\Delta\bar{Q}_{jCh} < \Delta\bar{Q}_{jCih}$ , which reduces the effective cross-section.  $\Delta\bar{Q}_{jCh}$  is related to the internal BC space charge density  $\bar{Q}_{jCi}$  according to 1D theory. At high current densities, the mobile charge in the BC depletion region associated with  $J_{nx}$  changes the lateral distribution of  $\Delta\bar{Q}_{jCh}$  so that an additional (2D related) bias dependence occurs.

The minority charge at low current densities consists mainly of electrons in the base region, which are confined vertically between the junctions. Laterally, there is some spread outside the emitter window that is mostly caused by perimeter injection. Thus, the neutral base minority charge density  $\Delta\bar{Q}_{mB}$  and its weighted counterpart,

$$\Delta\bar{Q}_{mBh} \cong q \int_{x_e}^{x_c} h_g \Delta n dx, \quad (5.100)$$

vary laterally at the emitter window edge and perimeter junction through both  $\Delta n$  and the integration limits. Taking again the weighted internal charge density  $\Delta\bar{Q}_{mBih} = \Delta\bar{Q}_{mBh}(y=0)$  as reference, assuming negligible lateral dependence under the emitter window, and relating  $\Delta\bar{Q}_{mBih}$  to the 1D base minority charge density  $\bar{Q}_{nB}$  gives according to (4.78)

$$\Delta\bar{Q}_{mBh} \cong \bar{h}_{mB2} A_{E0} \bar{Q}_{nB} \quad (5.101)$$

with the 2D weight factor

$$\bar{h}_{mB2} = \bar{h}_{mB} \frac{2}{b_{E0}} \left[ \int_0^{b_{E0}/2} h_{Jv,y} dy + \int_{b_{E0}/2}^{b_y} h_{Jv,y} \frac{\Delta\bar{Q}_{mBh}}{\Delta\bar{Q}_{mBih}} dy \right]. \quad (5.102)$$

The effective cross-section in the brackets is larger (smaller) than  $b_{E0}/2$  if the base minority charge density injected over the perimeter junction is larger (smaller) than the charge density under the emitter. This is typically the case if the base doping decreases (increases) towards the surface, but also depends on the material composition. Further examples can be found in [26, 27].

The discussion above results for each weighted charge component in a relation of the form  $\Delta\bar{Q}_{\kappa h} \cong \bar{h}_{\kappa 2} A_{E0} \bar{Q}_{\kappa}$ , in which the analytical expressions for the actual 1D charge densities can be inserted. Any bias depen-

dence of the corresponding effective cross-section is included in the weight factor  $\bar{h}_{k2}$ . As it turns out, the difference in bias dependence between 1D and 2D weight factors is fairly small [27].

A method for the experimental determination of an effective electrical emitter width for the collector and base current was discussed in [22]. Here, the extension of the emitter width is defined by the ratio of perimeter current per length,  $I_{T,P}$ , to the internal current density,  $I_{T,A} = J_{T,i}$  (cf. (5.68)) and can be easily calculated from device simulation results as

$$\gamma_C = \frac{I_{T,P}}{I_{T,A}} = \frac{I_C - J_C A_{E0}}{J_C A_{E0}}. \quad (5.103)$$

Here,  $I_C$  is the terminal current of 2D or 3D simulation while  $J_C$  is the terminal current density of the corresponding 1D simulation (or 2D emitter center line) at the *same* internal control voltage  $V_{B'E'}$ . As shown in Fig. 5.28 the results obtained for  $\gamma_C$  are similar to those calculated from (5.74) over the entire bias region. Since device simulation allows to access an internal voltage  $V_{B'E'}$ , the calculation (5.103) can be extended towards high injection. This is very difficult for experimental data though. Nevertheless, this method, which is easily extended to charges and capacitances, allows a first-order experimental determination of the effective electrical emitter width  $b_E$  entering the analytical transfer current expression via (5.94b). Generalizing this result to an effective electrical emitter *area*  $A_E = b_E l_E$  (cf. (5.66)) yields for the total transfer current

$$I_T = I_{T,A} A_E. \quad (5.104)$$

According to Fig. 5.28b HBTs exhibit a significantly larger current dependence of  $\Delta b_E$  than BJTs. On the other hand, the decrease of  $\Delta b_E$  and thus  $b_E$  in the transfer current expression is partially compensated for by a similar decrease of  $Q_{p0h}$  in the denominator of (5.94a). This is one reason why the empirical approach in [22] has been working so well for quite a long time. For both BJT and HBT the  $V_{B'C'}$  dependence is very small up to current densities beyond peak  $f_T$ . At very high current densities the tendency of the voltage dependence follows that of the transit time, which in turn is determined by the critical current  $I_{CK}$ .

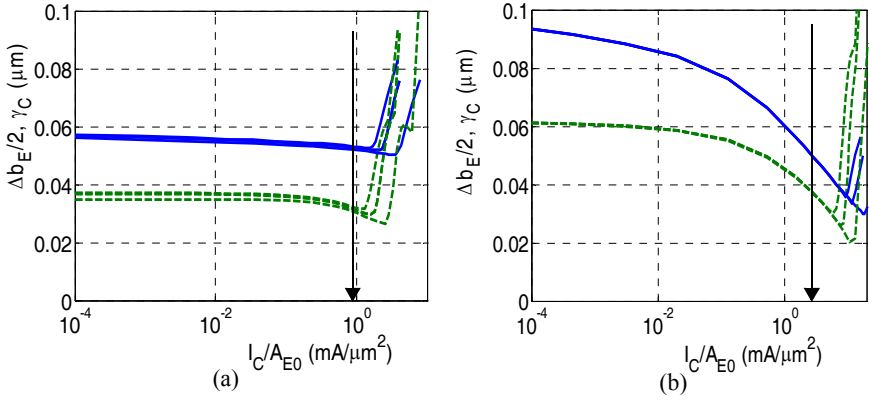


Fig. 5.28: Comparison between  $\Delta b_E/2$  from (5.74) (solid lines) and  $\gamma_C$  (dashed lines) as function of collector current density for  $V_{B'C}/V = 0.4, 0, -2.3$ : (a) BJT and (b) HBT. The arrow indicates peak  $f_T$ .

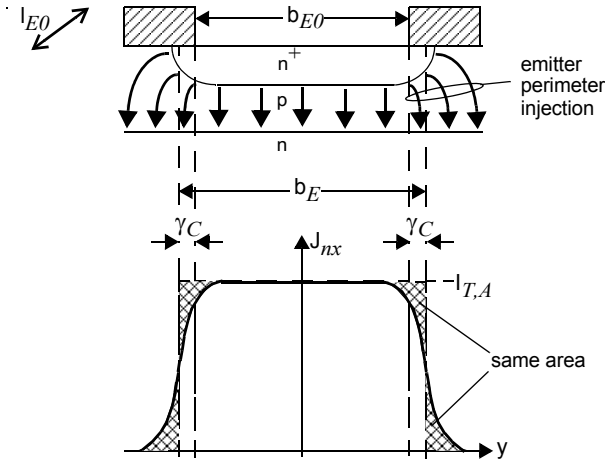


Fig. 5.29: Visualization of the physical meaning of the effective electrical emitter area. Schematic cross-section of the BE region with carrier injection paths (upper figure). Lateral distribution of injected electron current density at the BE junction (lower figure), and construction of the effective internal transistor with constant injection across the emitter width and zero perimeter current. The emitter length  $l_{E0}$  indicates the extension in perpendicular direction.

The effective emitter width corresponds to an effective electrical internal transistor containing the total transfer current with an average current density  $I_{T,A}$  given by (5.104) and a remaining zero perimeter injection. This is visualized in Fig. 5.29. The internal current density  $I_{T,A}$  is extended laterally until the total perimeter current fits under the box-like injection profile over the extended (i.e. effective) width.

### 5.3.3 Base current components

Figure 5.30 sketches various base current components that are associated with “regular” forward operation outside of the avalanche region. Under these conditions the main current flows into the internal transistor. The associated components have already been discussed in section 3.3. Similar components exist also in the emitter perimeter region. The backinjection across the emitter perimeter junction can be described by (cf. (3.101))

$$I_{jBEp} = I_{BEpS} \left[ \exp\left(\frac{V_{B^*E}}{m_{BEp} V_T}\right) - 1 \right] \quad (5.105)$$

with  $I_{BEpS}$  as the corresponding saturation current and  $m_{BEp}$  as the ideality coefficient. Note that also the control voltage has changed to the voltage  $V_{B^*E}$ , across the perimeter junction. Carrier recombination in the perimeter BE SCR can be described similarly to (3.113),

$$I_{REp} = I_{REpS} \left[ \exp\left(\frac{V_{B^*E}}{m_{REp} V_T}\right) - 1 \right], \quad (5.106)$$

with  $I_{REpS}$  and  $m_{REp}$ , respectively, as the corresponding saturation current and ideality coefficient, respectively. In addition, a current  $I_{BRs}$  may exist that represents the surface recombination at the interface to the spacer oxide. While this component is very important in III/V-HBTs it is generally negligible in Si-based transistors due to the “natural” property of the silicon-dioxide. Especially in SiGe HBTs even reverse bias stress has been shown to cause only little change in the BE related recombination current [28].

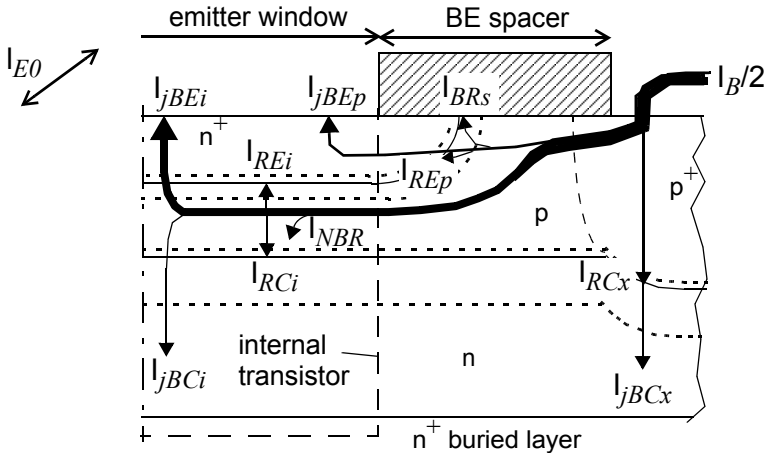


Fig. 5.30: Base current components related to the BE and BC regions. The dotted lines represent the SCR edges. For symmetry reasons, only half of a structure is shown.

At high forward operation, the barrier effect in HBTs leads to an additional recombination in the neutral base region that is represented by  $I_{NBR}$ . In contrast, the currents across the BC junction and their recombination portion are usually negligible even up to high current densities or low CE voltages. If necessary, these currents can be described by expressions similar to the above relations and generally need to be included in a compact model in order to flag a forward biased BC diode or substrate transistor. At very low CE terminal voltages the external component  $I_{jBCx}$  starts to supply the emitter current of the parasitic pnp transistor in a bulk process (cf. section. 5.4).

It should be noted that the spacer region has often also been called *link* region, especially for self-aligned poly-silicon BJTs. However, in HBTs with selective epitaxial growth the designation link usually refers to the narrow region connecting the poly-silicon with the mono-silicon base layer.

Geometry scaling of the components related to the BE junction can be accomplished in many processes by the standard-scaling approach. Applying (5.65) to, e.g., the emitter backinjection components yields according to (5.66) an associated effective electrical area  $A_{I_{jBE}}$  with the process-specific parameter

$$\gamma_B = \frac{I_{jBEp}/P_{E0}}{I_{jBEi}/A_{E0}} \cong \frac{I_{BEpS}/P_{E0}}{I_{BEiS}/A_{E0}}. \quad (5.107)$$

Since  $\gamma_B$  does not necessarily equal  $\gamma_C$  the corresponding effective areas are generally not the same. This has certain implications for the construction of a single-transistor equivalent circuit that are discussed later. Also, in contrast to the transfer current, the emission coefficients of the area and perimeter related base current components are generally different, e.g.  $m_{BEi} \neq m_{BEp}$ . As a consequence, using an effective area (e.g.  $A_{jBE}$ ) defined by the base current components is not very suitable for a compact model. Nevertheless, (5.65) is still required to separate the area and perimeter related portion of each base current component.

At higher CE voltages, impact ionization can occur in the BC SCR, which has already been discussed in section 3.3.4. As long as the electric field distribution is not influenced by the transfer current the main portion of the avalanche current is generated in the internal transistor while some contribution can also come from the external region. Applying the standard scaling approach yields for the total avalanche current

$$I_{AVL} = I_{AVL,A} A_{AVL} \quad (5.108)$$

with  $I_{AVL,A}$  as the avalanche current per area that can be described by the 1D analytical expression. If the internal and external collector doping are similar in the region containing  $I_T$ , the ratio of perimeter to area avalanche current is roughly the same as for the transfer current so that  $A_{AVL} = A_E$ . However, if a SIC exists that is implanted through the emitter window the electric field in the center region is larger than at the edge region. This results in a relative increase of the area contribution compared to the perimeter portion and thus  $A_{AVL} < A_E$ . Once at sufficiently high current densities the electric field distribution changes with current and starts to collapse in the center region, the area contribution decreases resulting in  $A_{AVL} > A_E$  with  $A_{AVL}$  now becoming bias dependent. In addition, the distributed avalanche current causes in the base region a voltage drop from the emitter center to the edge, leading to transfer current crowding in the center instead of at the edge. This so-called pinch-in effect, which can lead to destruction, is obviously of highly distributed nature and cannot be described at all by a lumped (i.e. single) transistor model [29]. Since a higher

temperature reduces impact ionization, electrothermal effects that lead to an increase in temperature in the center region compared to the external region can also have an influence on the location of the avalanche process. But this is only the case in transistors with long or wide emitters, i.e. if a significant temperature gradient exists already under the emitter. Then again a distributed model is required.

Tunneling is another mechanism that can lead to additional and usually undesired contributions to the base current. In most BJT technologies the tunneling current is dominated by its perimeter component due to the increase of base doping towards the surface under the spacer. This is not the case in (SiGe) HBTs in which tunneling occurs mainly under the emitter. For band-to-band tunneling (4.102) can be used with the appropriate field, bandgap, and voltage inserted for the area or the perimeter component. To describe the total tunneling current as a function of dimensions, the standard scaling approach can be applied, leading to

$$I_{btb} = I_{btb,A} A_{btb} \quad (5.109)$$

with  $I_{btb,A}$  as the 1D analytical description of the tunneling current (per area). The associated effective area  $A_{btb}$  does generally not equal  $A_E$ . For instance, a low base doping concentration at the perimeter junction leads to increased injection there but decreases tunneling. Also, the different field and bandgap results in different shapes of the area and perimeter related characteristics so that lumping these into a single component according to (5.109) may not always be accurate enough.

### 5.3.4 Charge storage

The emitter perimeter junction (cf. Fig. 5.18) not only contains an depletion charge but also leads to additional minority charge storage associated with the injected transfer and base current. Moreover, the perimeter transfer current has some influence on the BC depletion charge. The corresponding effects and options for geometry scaling are discussed below in more detail.

### 5.3.4.1 Depletion charges

In many processes, the standard scaling approach can be applied directly to the BE depletion charge and capacitance. Since the charge is measured via the capacitance, the latter is usually partitioned into its internal (bottom) portion  $C_{jEi}$  and its perimeter portion  $C_{jEp}$ . Applying (5.65) gives for the total BE depletion capacitance

$$C_{jE} = C_{jE,A}A_{E0} + C_{jE,P}P_{E0} \quad (5.110)$$

in which the specific components  $C_{jE,A}$  and  $C_{jE,P}$  are described by the analytical equations derived for the 1D case (cf. section 3.4) with the appropriate parameters and control voltages inserted. Using an effective emitter area according to (5.66) is possible, but requires usually a different set of model parameters compared to  $C_{jE,A}$  due to the difference in electrical behavior between internal and perimeter junction.

At first glance, the BC charge outside of the internal transistor (i.e. the external BC depletion charge) is not influenced by the emitter perimeter junction. However, the perimeter injection of a transfer current leads to a mobile carrier density also in the BC SCR under the spacer. Although this carrier density is smaller than under the emitter it can lead to a modulation of the electric field and, hence, a current dependent external BC depletion charge and capacitance. In a compact model, this increases the complexity of bias and geometry dependent calculations, since the ratio of current dependent to current independent portion changes with the dimensions of both emitter and external base. It is therefore desirable to lump the current dependent portion into the internal component using an effective internal collector area  $A_{jBCi}$  yielding

$$C_{jCi} = C_{jCi,A}A_{jBCi} \quad (5.111)$$

with  $C_{jCi,A}$  as voltage *and* current dependent 1D formulation.

Noticing that current spreading in the base and BC SCR is negligible,  $A_{jBCi} = A_E$  if the collector doping does not change under the spacer. This is the case if the SIC is implanted through the base poly opening. If the SIC is implanted through the emitter window one obtains  $A_{jBCi} < A_E$  at low current densities due to the drop of the capacitance per area outside of the emitter window. This also holds approximately at high current densities since the lower doping concentration causes the capacitance under the



spacer to become current dependent already at a lower current density than the capacitance under the emitter.

### 5.3.4.2 Minority charge

The perimeter minority charge is strongly linked to the injected perimeter current. Therefore, if the transfer current is lumped into a single component according to (5.65) and (5.104) the same should also apply to the minority charge associated with the transfer current. Strictly speaking, this does not include the emitter region related charge since this is coupled to the injected base current. However, the emitter perimeter minority charge is generally quite small so that as a first-order approximation the standard-scaling approach can be applied to the *total* minority charge

$$Q_m = Q_{m,A} A_{E0} + Q_{m,P} P_{E0}. \quad (5.112)$$

Figure 5.31 confirms the suitability of this expression based on device simulation. Here,  $Q_m$  was calculated directly from structures with different emitter widths and evaluated in the form

$$\frac{Q_m}{A_{E0}} = Q_{m,A} + Q_{m,P} \frac{P_{E0}}{A_{E0}}. \quad (5.113)$$

The result is indeed a straight line giving the correct 1D charge density  $Q_{m,A}$ , which was also calculated directly from a 1D structure and inserted on the y-axis. The slope gives the perimeter related specific charge  $Q_{m,P}$ .

Applying (5.113) to the whole bias region yields  $Q_{m,A}$  and  $Q_{m,P}$  as function of the terminal current  $I_C$  and voltage  $V_{BC}$ . However, each component should be dependent on its (area or perimeter) transfer current as well as on the respective internal and perimeter voltage in order to be able to apply the physics-based analytical expressions derived for the 1D case. Unfortunately, the separation of the area and perimeter current components is experimentally only possible for low injection (cf. section 5.3.1 for  $I_C$ ). Since it is not possible to determine  $Q_{m,A}$  and  $Q_{m,P}$  experimentally at higher current densities one has to describe the *total* (i.e. measurable) charge as a function of the *total* transfer current, *internal* BC voltage, and geometry.

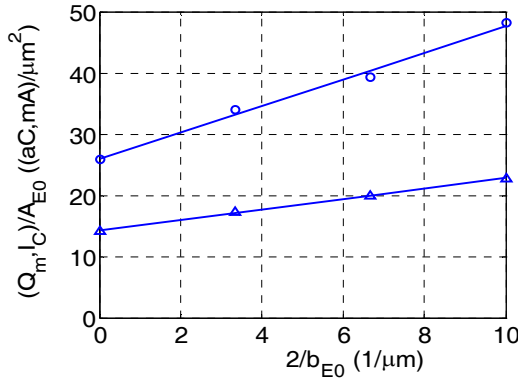


Fig. 5.31: Application of the standard-scaling approach (5.113) to the minority charge (circles) and collector current (triangles) at low current densities ( $V_{BE} = 0.75V$ ,  $V_{BC} = 0V$ ). The results were obtained from 2D device simulation of the HBT. The small “noise” in  $Q_m$  data is due to the numerical integration of the carriers.

At low injection,  $Q_{m,A}$  and  $Q_{m,P}$  are proportional to their corresponding current component via a minority storage time  $\tau_{m0,A}$  and  $\tau_{m0,P}$ , respectively, for the internal and perimeter region, respectively, leading to

$$Q_{m,A} = \tau_{m0,A} I_{T,A}, \quad Q_{m,P} = \tau_{m0,P} I_{T,P}. \quad (5.114)$$

With the specific transfer current components known from applying the standard-scaling approach, it is possible to determine the low-current minority storage times experimentally using (5.114). Together with (5.112), (5.103), and  $Q_m = \tau_{m0} I_T$  an explicit scaling equation for the total storage time at low current densities is obtained [25]:

$$\tau_{m0} = \tau_{m0,A} \frac{1 + \frac{\tau_{m0,P}}{\tau_{m0,A}} \gamma_C \frac{P_{E0}}{A_{E0}}}{1 + \gamma_C P_{E0}/A_{E0}}. \quad (5.115)$$

Figure 5.32a shows the geometry dependence of the normalized storage time for a fixed value of  $\gamma_C$  and selected values of  $\gamma_t = \tau_{m0,P}/\tau_{m0,A}$ . Figure 5.32b contains a magnified schematic view of a current flow line at the perimeter. Applying the existing theory to that flow line gives  $\tau_{m0,P} \propto w_B^2/F_z(w_B)$ . For most doping profiles the BC junction depth under the spacer either equals the one under the emitter or even increases so that  $\gamma_t > 1$ . This is expected to hold even if, like in SiGe HBTs, the base

doping decreases towards the surface since the resulting increase in mobility is compensated by a retarding drift field. According to (5.114) the influence of the perimeter charge storage increases with increasing  $\gamma_C$ . Only if the BC junction depth recedes with the emitter perimeter junction,  $\gamma_t$  may become  $\leq 1$  and the storage time decreases for narrow emitter widths. A special case is  $\gamma_t = \gamma_C = 0$  that corresponds to, e.g., a mesa-emitter structure without perimeter injection.

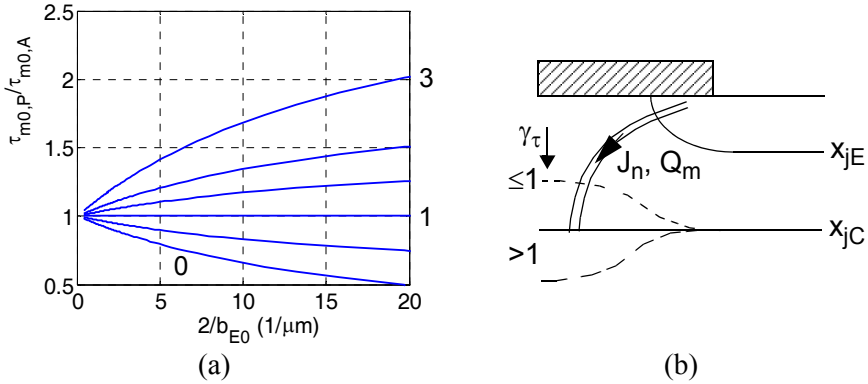


Fig. 5.32: (a) Geometry dependence of the normalized low-current storage time according to (5.114) with the ratio  $\gamma\tau = \tau_{m0,P}/\tau_{m0,A} = (0, 0.5, 1, 1.5, 2, 3)$  as parameter and  $\gamma_C = 0.052 \mu\text{m}$ . For the 2D case considered here  $P_{E0}/A_{E0} = 2/b_{E0}$ . (b) Schematic view of an infinitesimal current path element at the perimeter with different BC junction shapes.

Once at higher current densities the storage time becomes current dependent, the simple charge calculation needs to be replaced by  $Q_m = \int \tau_m dI$  making an explicit equation like (5.115) impossible. At high injection also the impact of collector current spreading needs to be taken into account, which will be discussed in section 5.4.

### 5.3.4.3 Base resistance under the BE spacer

The physical location of the spacer resistance  $R_{B_S}$  is shown in Fig. 5.33a. Its portion close to the emitter periphery behaves similarly in terms of bias due to the distributed current injection along the perimeter and the possible conductivity modulation of the sheet resistance  $r_{SB_S}(y)$  in that region. The

bias dependence decreases further away from the internal transistor. Finally,  $r_{SBs}(y)$  decreases rapidly towards the boundary to the poly-silicon region due to the outdiffusion there. The lateral variation of the base sheet resistance is displayed in Fig. 5.33b for the HBT. The first bias point is at low injection, the second just beyond peak  $f_T$ , and the third far into the high-current region. The behavior looks very similar over the whole base-collector voltage range.

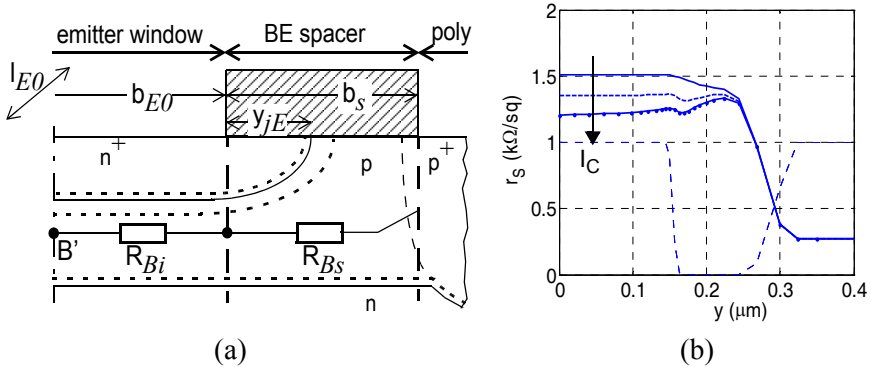


Fig. 5.33: (a) Location of spacer (or link) resistance component. The dotted lines represent the SCR edges. (b) Sheet resistance  $r_s$  of the HBT vs. lateral dimension for different bias points at  $V_{B'C'} = 0\text{V}$  and  $I_C/A_{EO} = (3 \cdot 10^{-3}, 5.14, 10.1) \text{ mA}/\mu\text{m}^2$ . The dashed line shows the normalized magnitude of the net doping at the surface  $x = 0$ .

According to the discussion above, the total resistance under the spacer can be partitioned into a bias dependent portion representing the region under the emitter perimeter junction and a bias independent portion representing the remaining region, giving for a structure with two base contacts

$$R_{Bs} = \frac{\overline{r_{SBs}} 2y_{jE}}{12 l_{EO}} \psi_s + \frac{1}{2l_{EO}} \int_{y_{jE}}^{b_s} r_{SBs}(y) dy. \quad (5.116)$$

Here,  $\overline{r_{SBs}}$  is a bias dependent *spatial* average value and  $\psi_s$  is the spacer current crowding function. Generally, current crowding in the short region under the perimeter junction is less pronounced than under the emitter.

Instead of trying to include a bias dependence in the spacer and, hence, external base resistance it seems more appealing to include the bias dependent spacer resistance portion into  $R_{Bi}$ . Considering the lateral variation of the sheet resistance in Fig. 5.33b a convenient approach is to include the

spacer region under the perimeter junction in the effective emitter width  $b_E$ . In other words, if  $b_{E0}$  in the expressions for  $R_{Bi}$  is simply replaced by  $b_E$  the bias dependent spacer portion can be included in the effective internal transistor, leaving just the remaining bias independent spacer component in the external base resistance. For negligible current crowding this approach corresponds to setting

$$\overline{r_{SBs}} 2y_{jE} \cong r_{SBi} \Delta b_E, \quad (5.117)$$

which appears to be justified according to Fig. 5.33b. At low (high) injection the extension leads to a slightly larger (lower) resistance than the actual spacer sheet resistance indicates. Note though that  $y_{jE}$  is only an approximate definition of the bias dependent spacer region boundary.

In BJTs the sheet resistance usually drops rapidly already within the spacer region under the emitter perimeter due to the high acceptor doping concentration at the surface. In this case, the extension results in a somewhat too large *bias dependent* base resistance component.

## 5.4 Three-dimensional and parasitic effects

The final step for making a compact model applicable to realistic structures is the extension of the formulations for the (effective) internal transistor and external regions to the 3D case. In the subsequent sections, first the general geometry scaling principle for junction related variables is extended to the 3D case. Then, the geometry dependence of the internal and external base resistance is considered for various practically relevant transistor configurations. Next, the elements of each external transistor region are discussed in terms of their geometry and possible bias dependence. The external transistor is defined such that it includes all regions that are invariant to the connection of a transistor in a circuit. In other words, regions located, e.g., above the silicon surface that are of parasitic nature but inherent to the structure itself need to be included in a model. This view is coherent with the set-up in process design kits in which the transistor model is supposed to include all elements of a p-cell that are associated with building the structure, while the parasitic extractor takes care of elements related to metal connections to other devices. Based on the results obtained in the

sections mentioned above, a complete physics-based equivalent circuit can then be constructed.

#### 5.4.1 Standard-scaling for a 3D structure

A fabricated pn junction on a chip has the mask dimensions  $b_0$  and  $l_0$  as shown in Fig. 5.34. The cross sections YY' and ZZ' along the y and z axis correspond to those discussed for the perimeter junctions so far, which have been described by the 2D standard-scaling approach (5.65). However, as a result of the dopant outdiffusion in *all* spatial directions lateral junctions also form at the four corners of the mask. Therefore, (5.65) has to be extended by the corresponding corner related component of (injected) current or charge, leading to the general expression

$$Y_e = Y_{e,A}A_0 + Y_{e,P}P_0 + 4Y_{e,C} \quad (5.118)$$

with  $Y_{e,C}$  as the contribution of a single corner junction to the variable under consideration. For determining the corner component experimentally usually structures with minimum mask (window) dimensions are required that are often difficult to fabricate reliably and to measure accurately.

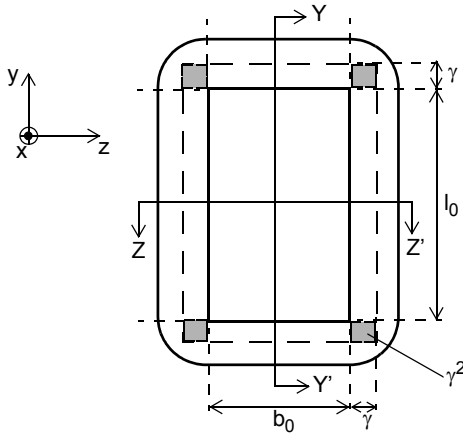


Fig. 5.34: Top view of a 3D pn junction and relevant dimensions.

If (5.65) is applicable to the 2D case, then inserting the resulting effective electrical dimensions  $b = b_0 + 2\gamma$  and  $l = l_0 + 2\gamma$  with  $\gamma$  defined by (5.68) leads to [25]

$$Y_e = Y_{e,A}A_0 + Y_{e,P}P_0 + 4Y_{e,A}\gamma^2. \quad (5.119)$$

This allows to express the corner contribution by already known quantities

$$Y_{e,C} = Y_{e,A}\gamma^2 = \gamma Y_{e,P}. \quad (5.120)$$

Lumping the last expression together with the perimeter component yields

$$Y_e = Y_{e,A}A_0 + Y_{e,P}P \quad (5.121)$$

with  $P = 2(b+l) = P_0 + 4\gamma$  as effective perimeter length. As a consequence, the existing 2D results can be used if just  $P_0$  is replaced by  $P$ . Moreover, the corner component does not have to be measured.

Above result is based on the assumptions that (i) the perimeter junction is the same in y and z direction, (ii) there is no “anomalous” diffusion at the corners, and (iii) the electrical junction behavior at the corner is the same as in y and z direction. While (i) and (ii) are usually fulfilled if (5.65) applies (iii) may be impacted by the somewhat different space-charge region width due to the spherically shaped junction. However, (5.120) has proved to be accurate according to 1D, 2D and 3D device simulation and also experimentally for many processes. However, in very advanced processes sometimes deviations from (5.118) or (5.120) are observed due to layout dependent profile changes (cf. section 5.5).

#### 5.4.2 Minority charge

As mentioned before and demonstrated in [25,24], collector current spreading can have a profound impact especially on charge storage. This can be observed directly in the transit time. For the comparison in Fig. 5.35 the collector current has been normalized to the effective emitter area defined by the 2D-GICCR. As a consequence, the currents densities of the 1D and 2D structure are aligned up to medium injection, making the weaker increase of the 2D curves clearly visible. This behavior is caused by collector current spreading. It is more pronounced for (older) BJTs than for SiGe HBTs.

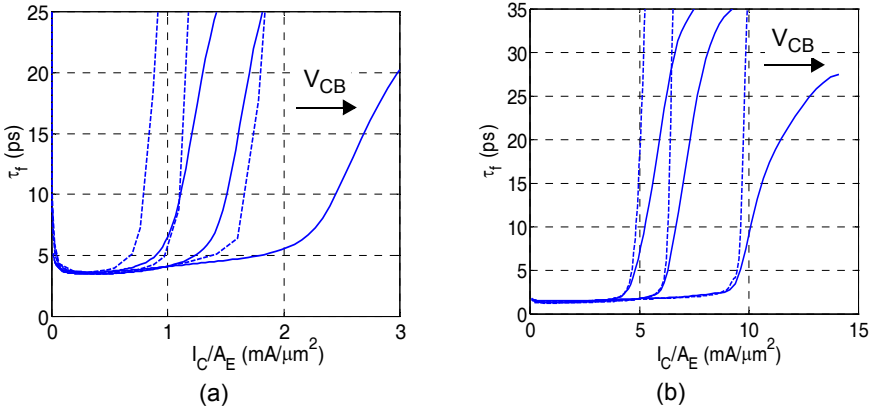


Fig. 5.35: Transit time, normalized to its low-current (minimum) value, vs. collector current density. Comparison between 2D device simulation (solid lines) and 1D device simulation (dashed lines): (a) BJT and (b) HBT.  $V_{BC}/V = 0.4, 0, -2.3$ .

As observed in Fig. 5.20 current spreading occurs already at low current densities in the non-depleted collector region, but it has no impact as long as the voltage drop there is negligible. With increasing current this voltage drop can become important at a sufficiently low voltage  $V_{ci}$  (cf. section 3.2, 4.1.3). In contrast to the 1D case, current spreading leads now to a smaller current density and, hence, also a smaller voltage drop in the ohmic region. For a given value of  $V_{ci}$  this results in a higher voltage and field across the depleted junction region. As a consequence, current spreading leads to a shift of the onset of high-current effects towards higher collector current densities. This shift becomes less pronounced at higher voltages  $V_{ci}$  due to the small (or zero) width of the non-depleted ohmic collector region.

According to section 3.5.2 the onset of high-current effects can be characterized by the critical current  $I_{CK}$ . Its low-voltage component  $I_{CKl}$  is defined by the condition of an entirely ohmic collector. This condition is illustrated for the 2D case in Fig. 5.36, assuming a current spreading angle  $\delta_C$  and an initial width  $b_E$  at the BC junction. The cross-section width in the collector then increases with depth  $x'$  according to

$$A_C = A_E(1 + \xi_b x') \quad (5.122)$$

with



$$\xi_b = \frac{2}{b_E} \tan(\delta_C) \quad (5.123)$$

as geometry dependent factor determined by the current spreading angle.

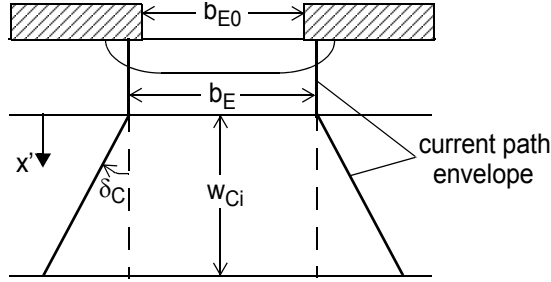


Fig. 5.36: Sketch of the envelope of the transfer current path and geometric model with the relevant variables used for deriving the collector current spreading function.

From the neutrality condition with  $I_{CK}$  as pure drift current follows

$$I_{CKI} = qN_{Ci}A_C(x')v(E_x). \quad (5.124)$$

In the 1D case, the vertical electric field  $E_x$  in the collector is constant ( $= V_{Ci}/w_{Ci}$ ). However, in the 2D and 3D case,  $E_x$  decreases with  $x$  due to current spreading and needs to be calculated first in order to obtain an analytical expression for  $I_{CK}$ . Inserting (2.11) with  $grad(\varphi_n) = E_x$  yields

$$I_{CKI} = I_{lim} (1 + \xi_b x') \frac{E_x(x')/E_{lim,n}}{\left[ 1 + \left( \frac{E_x(x')}{E_{lim,n}} \right)^{\beta_{\mu n}} \right]^{1/\beta_{\mu n}}}. \quad (5.125)$$

with  $I_{lim}$  defined by (3.4). While this equation can be solved for  $E_x$  the subsequent integration over  $x$  yields an implicit equation for  $V_{Ci}$  (or  $I_{CKI}$ ) the form of which also depends on  $\beta_{mn}$ . Such a solution, although exact if solved numerically, is not suitable for compact modeling. Therefore, the field in the denominator term in (5.125) is approximated by an average value  $\bar{E}_x$  leading to

$$E_x(x') = \frac{I_{CKI}}{I_{lim}} \left[ 1 + \left( \frac{\bar{E}_x}{E_{lim,n}} \right)^{\beta_{\mu n}} \right]^{1/\beta_{\mu n}} \frac{E_{lim,n}}{1 + \xi_b x'}. \quad (5.126)$$

Integration over the whole collector region  $[0, w_{Ci}]$  yields the internal (normalized) collector voltage

$$\frac{V_{ci}}{V_{lim}} = \frac{I_{CKI}}{I_{lim}} \left[ 1 + \left( \frac{\bar{E}_x}{E_{lim,n}} \right)^{\beta_{\mu n}} \right]^{1/\beta_{\mu n}} \frac{\ln(1 + \xi_b w_{Ci})}{\xi_b w_{Ci}}. \quad (5.127)$$

This leads to an explicit expression for the critical current

$$I_{CKI} = \frac{V_{ci}}{r_{Ci0}} \frac{1}{[1 + (V_{ci}/V_{lim})^{\beta_{\mu n}}]^{1/\beta_{\mu n}}} \frac{\xi_b w_{Ci}}{\ln(1 + \xi_b w_{Ci})}, \quad (5.128)$$

where (3.243) has been used. A comparison of this simple solution with the exact solution for  $I_{CKI}$  (or  $V_{ci}$ ) based on (5.125) showed that replacing  $\bar{E}_x$  by  $V_{ci}/w_{Ci}$  leads to an error of about 5% at  $V_{ci} = 2V_{lim}$ , which is already outside of the low-voltage range. The result for  $I_{CKI}$  differs from the 1D relation (3.249) only by the last term, which can be identified as a collector current spreading function.

For the 3D case, the cross-sectional area for the current density in the collector is given by

$$A_C = A_E(1 + \xi_b x')(1 + \xi_l x') \quad (5.129)$$

with

$$\xi_l = \frac{2}{l_E} \tan(\delta_C). \quad (5.130)$$

Extension of the theory above to the 3D case then leads to the collector current spreading function [25]

$$f_{cs} = \begin{cases} \frac{\zeta_b - \zeta_l}{\ln[(1 + \zeta_b)/(1 + \zeta_l)]} & , l_E > b_E \\ 1 + \zeta_b & , l_E = b_E \end{cases}. \quad (5.131)$$

with the width and length related spreading factors

$$\zeta_b = \xi_b w_{Ci} \quad \text{and} \quad \zeta_l = \xi_l w_{Ci} \quad (5.132)$$

introduced for convenience. The expression for the special case of a square emitter is obtained from a series expansion of the general case  $l_E > b_E$ . Letting  $l_E \rightarrow \infty$  gives the result consistent with the 2D case in (5.128).

The behavior of the current spreading function is visualized in Fig. 5.37a with dependence on the reciprocal spreading factors. This diagram allows for any combination (within a reasonable range) of technology parameters  $b_E$ ,  $l_E$ ,  $w_{Ci}$  to determine graphically the value of the collector spreading function for a given  $\delta_C$ . The cases below the dashed line, which corresponds to the square emitter function in (5.131), represent  $l_E > b_E$  and are usually of interest. For instance, at  $\zeta_b = 1$  moving up from the 2D case gives an increasing value of  $f_{CS}$  with decreasing emitter length of (10, 5, ... 1) times  $b_E$  due to the 3D contribution from the emitter foreside.

The impact of collector current spreading at low-voltages on the critical current is illustrated in Fig. 5.37b for selected cases. The lowest curve represents the 1D case. A finite emitter width (2D case) can increase  $I_{CKI}$  significantly. A finite length leads to further increases with the square emitter as the limiting case.

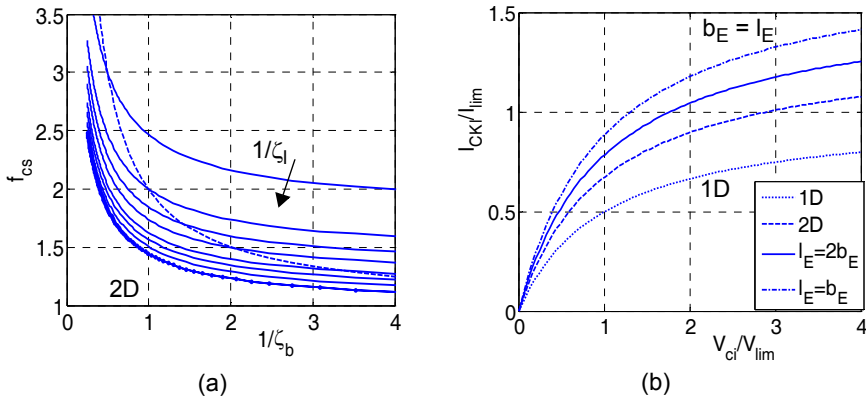


Fig. 5.37: (a) Collector current spreading function vs. reciprocal width related spreading factor,  $1/\zeta_b$ , with the parameter  $1/\zeta_l = 0.5, 0.7, 1, 2, 3, 5, 10, \infty$ . The dashed line represents square emitters and the solid line with dots the 2D case. (b) Low-voltage component of the normalized critical current vs. normalized internal collector voltage for different emitter geometries (parameters:  $\delta_C = 30^\circ$ ,  $\beta\mu\nu = 1$ ,  $b_E = 0.3\mu\text{m}$ ,  $w_{Ci} = 0.2\mu\text{m}$ ).

Collector current spreading has a smaller impact on  $I_{CK}$  at high voltages, since the field in the remaining collector region is still very high. Only at the end of the neutral base some current spreading can occur at  $I_{CK}$  due to the reduced field at the BC junction. For modeling the 2D and 3D case the

high-voltage expression  $I_{CKh}$  from (3.250) has to be merged with the low-voltage expression  $I_{CKlf_{cs}}$ .

At current densities beyond  $I_{CK}$  an injection zone forms. A fairly detailed derivation of the stored minority charge in the injection zone was given in [25] and shall not be repeated here. Below, only the main assumptions will be discussed and the results will be summarized.

First, the reasonable assumption is made that the voltage drop across the injection zone is negligible and, thus, the entire voltage  $V_{ci}$  drops across the remaining collector region. The corresponding situation is sketched in Fig. 5.38. On the right the schematic dependence of the electric field is shown for a low and a high voltage. Due to the low field in the injection zone (with width  $w_i$ ) current spreading occurs in this region with the envelope of the current path sketched on the left. In the remainder of the collector the current spreading continues for low voltages while at high voltages the current is focused vertically again due to the high electric field there.

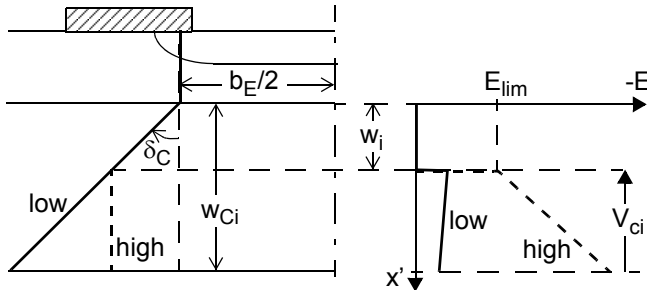


Fig. 5.38: Collector current spreading at high current densities: sketch of the geometric model (left) and of the electric field (right) at a low and a high voltage.

From the assumption above the voltage drop  $V_{ci}$  can be calculated similarly to (5.127) but with  $I_{CK}$  replaced by  $I_C$  and the integration now being performed from  $w_i$  to  $w_{Ci}$ . This allows to calculate the normalized injection width  $w_{in} = w_i/w_{Ci}$  for the 3D case as a function of process parameters,

$$w_{in} = \begin{cases} \frac{\kappa - 1}{\zeta_l - \kappa \zeta_b} & , l_E > b_E \\ \frac{1}{\zeta_b} \left[ \frac{1 + \zeta_b}{\zeta_b(I_{CK}/I_C) + 1} - 1 \right] & , l_E = b_E \end{cases} , V_{ci} < V_{lim} \quad (5.133)$$

with

$$\kappa = \frac{1 + \zeta_l}{1 + \zeta_b} \exp \left[ \frac{I_{CK}}{I_C} \ln \left( \frac{1 + \zeta_b}{1 + \zeta_l} \right) \right]. \quad (5.134)$$

At high voltages the current path in the SCR at the end of the collector is one-dimensional but with a larger cross section that depends on  $w_i$ . Integrating the corresponding Poisson equation twice leads to a cubic equation for  $w_i$  as function of  $V_{ci}$  and  $I_C$  (cf. [25]). This is not very attractive for compact modeling since the (numerical) solution of this equation still needs to be coupled smoothly to the low voltage expression.

The stored minority charge in the collector can be calculated regardless of the expression of the injection width, i.e. with having to distinguish between low and high voltages. As in the 1D case, the derivation starts with (3.267). Using the conclusion obtained there that the hole density  $p$  decreases linearly with depth and formulating (3.267) for a spatially dependent current density  $J_{nx}$  yields for the 3D case and within the interval  $0 \leq x \leq w_i$ :

$$I_T = -2q\mu_{nC0}V_T A_C(x') \frac{dp}{dx'} \quad (5.135)$$

with the effective collector area from (5.129). Integrating twice over the injection region yields the associated charge

$$Q_{pC} = \tau_{pCs} I_T \begin{cases} 2 \frac{f_1 \ln \left( \frac{1 + \zeta_b w_{in}}{1 + \zeta_l w_{in}} \right) - f_2 - f_3}{\zeta_b - \zeta_l} & , l_E > b_E \\ \frac{1 + \zeta_b w_{in}/3}{1 + \zeta_b w_{in}} w_{in}^2 & , l_E = b_E \end{cases} \quad (5.136)$$

with  $\tau_{pCs}$  from (3.274) and the auxiliary functions

$$f_1 = w_{in} + \frac{\zeta_b + \zeta_l}{2} w_{in}^2 + \frac{\zeta_b \zeta_l}{3} w_{in}^3, \quad (5.137a)$$

$$f_2 = \frac{1 - \frac{\zeta_b}{\zeta_l}}{\zeta_b} \left[ (1 + \zeta_b w_{in})^2 \left( \frac{\ln(1 + \zeta_b w_{in})}{2} - \frac{1}{4} \right) + \frac{1}{4} \right] \quad (5.137b)$$

$$+ \frac{1}{\zeta_b \zeta_l} \left[ (1 + \zeta_b w_{in})^3 \left( \frac{\ln(1 + \zeta_b w_{in})}{3} - \frac{1}{9} \right) + \frac{1}{9} \right]$$

and  $f_3$  having the same form as  $f_2$  but with  $\zeta_b$  and  $\zeta_l$  interchanged.

The above result is quite elaborate but has proved to enable accurate modeling of 2D and 3D effects. In modern high-speed transistors, the selectively implanted collector (SIC) doping has increased significantly beyond the background doping of the external collector. The lateral SIC doping decrease towards the external transistor results in a fairly high lateral electric field pointing towards the internal transistor and, hence, focuses the transfer current vertically. This reduces collector current spreading at high injection, especially for implants through the emitter window. Furthermore, the BC barrier in SiGe HBTs reduces also the relevance of collector minority charge storage. Therefore, in advanced SiGe HBTs it is often sufficient to consider only the current spreading factor in the critical current, but to ignore the  $Q_{pC}$  related calculations. However, since  $Q_{pC}$  also enters the transfer current formulation via the 2D/3D-GICCR with a weight factor that can be much larger than 1 it may be necessary to include current spreading in the weighted collector charge. For this a similar derivation as for (5.136) can be performed with the current spreading in the weight factor  $h_f(x,y)$  described by (5.129).

### 5.4.3 Base resistance

Figure 5.39a visualizes the various base regions in a transistor with a single base contact on one side along the emitter. The top figure shows a schematic cross-section as reference for a lateral cut through the base region. This  $y$ - $z$  cut defines a 2D plane containing the (mostly) lateral current

flow with the relevant components indicated by the bold arrows. The current injected into the emitter is indicated by the arrows pointing upwards in -x-direction. According to previous discussions, the effective emitter dimensions are used. Figure 5.39b shows the base resistance components associated with the different physical regions. The internal base resistance  $R_{Bi}$  has already been discussed, but only for the 2D case. Its formulation will be extended here to the 3D case. Then, each of the other components will be considered for various base and emitter contact arrangements, resulting in a geometry scalable compact formulation for the total external and internal base resistance.

Following the same definition (5.18) as for  $R_{Bi}$ , the total DC base resistance may be defined as

$$R_B = \frac{V_{BE} - V_{B'E'}}{I_B} \quad (5.138)$$

with  $I_B$  as total base current at the base terminal and  $V_{BE}$  as the BE terminal voltage. In order to be able to determine  $R_B$  from a measured  $I_B(V_{BE})$  characteristic, the relation between  $I_B$  and the internal BE voltage  $V_{B'E'}$  must be known. For a compact model the most convenient representation of  $R_B$  is a partitioning into  $R_{Bi}$  and a lumped external component  $R_{Bx}$ . Then, with  $I_B = I_{Bi} + I_{Bp}$  (5.138) reads

$$V_{BE} - V_{B'E'} = R_{Bx}I_B + R_{Bi}I_{Bi} = R_{Bx}I_{Bp} + R_{Bi}I_{Bi}. \quad (5.139)$$

Assuming a constant ratio  $g_B = I_{Bp}/I_{Bi}$  allows to write

$$V_{BE} - V_{B'E'} = (R_{Bx}g_B + R_{Bi})I_{Bi}, \quad (5.140)$$

in which  $I_{Bi} = I_{BEi}$  provides the desired relation to  $V_{B'E'}$ . Hence, the conventional method for determining (and verifying)  $R_B$  from a simple  $I_B(V_{BE})$  measurement does not yield  $R_B$  but a modified resistance that depends on the perimeter injection. Furthermore, at high forward bias, where the deviation of  $I_B(V_{BE})$  from an ideal characteristic is most detectable,  $g_B$  cannot be assumed to be constant due to the different injection behavior of  $I_{Bi}$  and  $I_{Bp}$ . Moreover, the BC barrier effect in SiGe HBTs makes the  $I_{Bi}(V_{B'E'})$  characteristic non-ideal. As a consequence, (5.138) is generally not suitable for an accurate determination of the base resistance. Also, it

would not provide any information about the physical components of  $R_B$  which is necessary for a geometry scalable model.

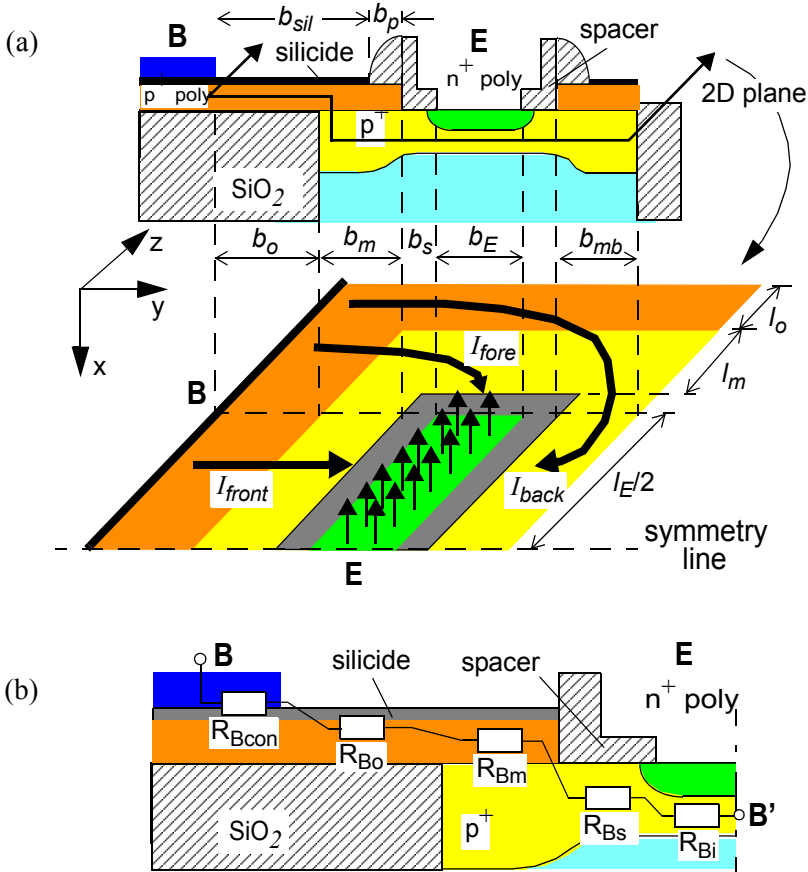


Fig. 5.39: (a) Schematic cross-section (top) of a transistor with single base contact and projection of lateral base current flow into a 2D plane (bottom). The long bold arrows in the yz-plane indicate the relevant current contributions to the front, fore and back side of the structure, while the short arrows pointing out of the internal base region (in -x-direction) represent the base current flow across the BE junction. (b) Magnified view of the base regions with the corresponding resistance components.

Using device simulation it is possible though to eliminate the issues mentioned above, which are caused by effects associated with the vertical current flow, while keeping the relevant resistive current flow in lateral direction. For this, instead of a full 3D transistor structure just a 2D plane



through the base is simulated as indicated in Fig. 5.39a, which contains the projection of the flow lines associated with the base current. The main current components that support the current flowing into the internal base and their basic flow are indicated by  $I_{front}$ ,  $I_{fore}$ , and  $I_{back}$ .

The calculation of  $R_B$  for general emitter geometries is based on the solution of the hole transport and continuity equation (cf. chapter 2) in the 2D plane shown in Fig. 5.39a. Integration over the relevant vertical base region leads for the 2D transport equation to the per length current vector in the  $(y, z)$ -plane,

$$\vec{I}_p = (I_{py}, I_{pz}) = \int (J_{py}, J_{pz}) dx = -(q \int p \mu_p dx) (\text{grad} \phi_p), \quad (5.141)$$

with  $I_{py}$  and  $I_{pz}$  as averages over the vertical dimension. Due to the high base doping and small vertical hole current the potential  $\phi_p$  does not depend on  $x$  over the interval  $[x_e, x_c]$ , where the hole density significantly contributes to the integral and therefore equals the voltage across the junction (with the emitter being grounded). Recognizing the first portion of the rightmost equation as the (reciprocal of the) internal base sheet resistance (cf. (5.28)) simply leads to

$$r_{Sbi} \vec{I}_p = -\text{grad} \phi_p. \quad (5.142)$$

Also, with  $\phi_p$  as the controlling voltage the vertical current density component at the emitter edge  $x_e$  of the neutral base represents the backinjection and reads

$$J_{px}(x_e) = -J_{BEi} = -J_{Sbi} \exp\left(\frac{\phi_p(y, z)}{V_T}\right). \quad (5.143)$$

Neglecting base volume recombination in the hole continuity equation and moving the vertical current density component  $J_{px}$  to the right-hand-side yields after integration in vertical direction

$$\text{div} \vec{I}_p = \int \left( \frac{dJ_{py}}{dy} + \frac{dJ_{pz}}{dz} \right) dx = - \left( \int_{J_{px}(x_e)}^{J_{px}(x_c)} dJ_{px} + \frac{\partial(q \int p dx)}{\partial t} \right). \quad (5.144)$$

Recognizing  $\bar{Q}_p = q \int p dx$  as the hole charge per area one obtains

$$\text{div} \vec{I}_p = -[J_{px}(x_c) - J_{px}(x_e)] - \frac{\partial \bar{Q}_p}{\partial t}. \quad (5.145)$$

Since for practical applications  $J_{px}(x_c) \ll J_{px}(x_e)$  one obtains with (5.143)

$$\text{div} \vec{I}_p = - \left( J_{SBI} \exp \left( \frac{\varphi_p}{V_T} \right) + \frac{\partial \bar{Q}_p(\varphi_p)}{\partial t} \right). \quad (5.146)$$

Compared to the corresponding equations, e.g., (5.5) and (5.6), used in the previous 2D analyses the “-” sign here and in (5.142) results from the different definition of  $\vec{I}_p$  w.r.t. to the y-axis.

Eqs. (5.142) and (5.146) are a quasi-3D representation of the emitter current crowding problem, which can be solved for arbitrary emitter geometry using 2D device simulation (e.g. [30][31][32][33]). For this, the usual recombination expression in a device simulator is replaced by  $J_{BEI}$  according to (5.143). For calculating the geometry dependence of the DC base resistance the exact values of the sheet resistances (including  $r_{SBI}$ ) and dimensions in Fig. 5.39 are not important as long as realistic *ratios* are assumed. In this case the sheet resistance is simply expressed as

$$r_{SBI} = (q \bar{\mu}_p \bar{N}_B L_x)^{-1}. \quad (5.147)$$

The average values for mobility and doping are adjusted to the desired resistance value. The bias dependence of the vertical components  $r_{SBI}$  and  $\bar{Q}_p$  can be described analytically or provided through a table.

For investigating the base resistance under small-signal operation, the time derivative of the hole charge density is transformed into the frequency-domain,

$$\frac{\partial \bar{Q}_p}{\partial t} = \frac{\partial \bar{Q}_p}{\partial \varphi_p} \frac{\partial \varphi_p}{\partial t} \rightarrow j\omega \bar{C}_B \varphi_p, \quad (5.148)$$

with  $\bar{C}_B$  as the sum of the capacitances per area at each point ( $y, z$ ) and  $\varphi_p$  as the complex valued potential. In a device simulator,  $\bar{C}_B$  can be described analytically as a function of bias and replaces  $p$  in the corresponding code portion.

The special device simulation described above has been used to calculate the geometry and bias dependence of the internal, external, and total base resistance for a variety of emitter sizes and contact configurations in order to provide a reference for analytical solutions of special cases and for approximations in general cases (e.g. [34][30][31][32][35]). Below, first

the 2D description of the internal base resistance is extended to the 3D case. Then, a compact geometry dependent analytical description is derived for the external and total base resistance.

#### 5.4.3.1 Internal base resistance for the 3D case

##### A Static operation

Combining (5.142) with (5.146) and dropping the time derivative leads to a nonlinear second-order differential equation for an arbitrary emitter geometry,

$$\text{div}(\text{grad}\phi_p) = r_{SBI} J_{SBI} \exp(\phi_p/V_T), \quad (5.149)$$

in which  $r_{SBI}$  is assumed to be spatially independent based on the same considerations regarding conductivity modulation as in section 5.2.1.2. Even under this assumption, closed-form analytical expressions for the base resistance can only be obtained for special emitter geometries and certain operating conditions. Results for important special cases are given below before a model for general rectangular emitter geometries is discussed.

For a circular emitter, (5.149) can be simplified by a coordinate transformation resulting in the radius as a single spatial variable. Assuming a surrounding contact as boundary condition the corresponding low-injection solution for the internal base resistance was given first in [36],

$$R_{Bi,nc}^{circ} = \frac{r_{SBI}}{8\pi}, \quad (5.150)$$

but without any derivation. The current dependence was first derived in [16], yielding the current crowding function

$$\psi(\eta) = \frac{\ln(1 + \eta)}{\eta} \quad (5.151)$$

and the current crowding factor

$$\eta = \frac{r_{SBI} I_{Bi}}{8\pi V_T}. \quad (5.152)$$

For rectangular emitters with a surrounding contact an exact solution for arbitrary aspect ratios can be obtained for negligible current crowding, i.e. when the back injection current density  $J_{BEi}$  is homogeneous across the emitter window, from a Fourier series [35]:

$$R_{Bi,nc} = r_{SBI} \frac{b_E}{l_E} \frac{2}{\pi^2} \sum_{\lambda \text{ odd}} \frac{1}{\lambda^2} \frac{\lambda \alpha - \tanh(\lambda \alpha)}{(\lambda \alpha)^3} \quad (5.153)$$

with

$$\alpha = \frac{\pi b_E}{2 l_E}. \quad (5.154)$$

A closed-form solution for the current crowding function is unknown, however, for the case of an arbitrary aspect ratio.

One of the extreme cases is the square emitter, which - according to Fig. 5.40a - exhibits a symmetrical 2D current flow in the  $yz$ -plane. The low-current value can be obtained directly from (5.153),

$$R_{Bi,nc} = r_{SBI} \sum_{\lambda \text{ odd}} \frac{16}{(\lambda \pi)^5} \left[ \lambda \frac{\pi}{2} - \tanh\left(\lambda \frac{\pi}{2}\right) \right] = \frac{r_{SBI}}{28.45}. \quad (5.155)$$

In order to be able to provide a complete geometry scalable internal base resistance description the current crowding function for this case is also of interest. By neglecting the corner related current crowding the same form as (5.151) and (5.152) for the circular emitter was obtained in [34], except for the factor 32 instead of  $8\pi$  ( $= 25.1$ ). Detailed quasi-3D simulations [30] [31], which included corner current crowding, resulted in the factor 28.6 which is close to the theoretical value 28.45 in (5.155) derived later.

The computationally expensive expression (5.153) cannot be easily extended to include corner rounding through, e.g., the normalized rounding radius  $r' = b_{Er}/(b_E/2)$ . Therefore, the expression from [30] is extended to combine all results above into a simple single geometry function

$$g_i(b_E, l_E) = \frac{1}{12} + \left( \frac{1-r'}{28.45} + \frac{r'}{8\pi} - \frac{1}{12} \right) \frac{b_E}{l_E}. \quad (5.156)$$

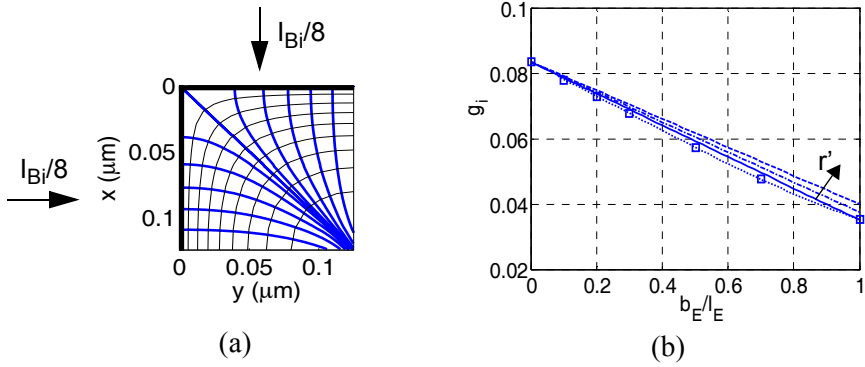


Fig. 5.40: (a) Current flow and equipotential lines for the internal transistor obtained from quasi-3D simulation for a square-emitter with surrounding contact; only one quarter of a complete structure is shown. (b) Geometry function (5.156) vs. emitter width to length ratio for selected normalized corner radii  $r' = 0, 0.5, 1$ . Simulated values (symbols) for a rectangular emitter are inserted as reference, and the exact analytical geometry function (5.153) is inserted as comparison (dotted line).

The low-current internal base resistance then reads

$$R_{Bi, nc} = r_{SBi} \frac{b_E}{l_E} g_i(b_E, l_E). \quad (5.157)$$

Figure 5.40b compares the geometry dependence of  $g_i$  from (5.156) with the exact solution from (5.153) for rectangular structures and the corresponding results from device simulation. The deviations of (5.156) are certainly much smaller than process variations of  $r_{SBi}$  and dimensions. Figure 5.40b also contains  $g_i$  from (5.156) for different values of  $r'$ .

For higher current densities also a unified DC current crowding function is required. Following [30], the circular emitter crowding function (5.151) is selected since it is the most simple expression and is also obtained for square emitters [34]. As a consequence, the current crowding factor (5.15) is generalized to

$$\eta = \frac{r_{SBi}}{g_\eta(b_E, l_E)} \frac{b_E I_{Bi}}{l_E V_T} = \frac{R_{Bi, nc} I_{Bi}}{V_T} \frac{1}{g_\eta g_i}, \quad (5.158a)$$

where the geometry dependent function

$$g_{\eta} = 19.2 \left( \frac{b_E}{l_E} \right)^2 - 11.9 \frac{b_E}{l_E} + 17.84 \quad (5.158b)$$

makes (5.151) valid for different aspect ratios [30]. The parameter values are determined in a first step from making the simulated results of rectangular structures with different aspect ratios to overlay each other. In a second step the slight difference between this unified current dependence and (5.151) is eliminated by multiplying the first parameter set with  $8\pi/28.45$  that is close to 0.9 found empirically in [30]. Hence, this multiplication leads to  $g_h(b_E/l_E=1) = 8\pi$ . This is justified, since according to Fig. 5.40a the equipotential lines in a square structure already start being rounded very close to the corners despite the fixed potential there on the edge. This is much more pronounced in realistic structures with external regions (see later) and with rounded emitter windows for lithography reasons, in which the equipotential lines resemble those in a circular structure fairly well. Overall, (5.158a,b) ensure that (5.151) gives the same result as (5.26) down to at least  $\psi(\eta) = 0.5$ .

## B Small-signal frequency dependent operation

From (5.142) and (5.146) follows after superimposing on  $\varphi_p$  a small-signal variation  $v(t) \ll \varphi_p$ , series expansion of the exponential function, separation of the resulting PDE into a DC and AC part, and subsequent transformation of the AC equation into frequency domain

$$r_{Sbi} \vec{I}_p = -\text{grad} \underline{V}, \quad (5.159)$$

$$\text{div} \vec{I}_p = -(\bar{G} + j\omega \bar{C}_B) \underline{V}. \quad (5.160)$$

with  $\underline{V}(\omega)$  as frequency domain representation of  $v(t)$ . Combining the two equations yields

$$\text{div}(\text{grad} \underline{V}) = \underline{k}^2 \underline{V} \quad (5.161)$$

with the complex frequency dependent variable

$$\underline{k} = \sqrt{r_{Sbi}(\bar{G} + j\omega \bar{C}_B)} = \sqrt{r_{Sbi}} \underline{\bar{Y}}. \quad (5.162)$$

For the general case of a spatially dependent  $k(y,z)$ , i.e. DC current crowding, the solution can only be obtained by numerical means (e.g. [3][9][10]), even for the 2D case. Closed-form solutions can be found only for special cases if  $k$  is spatially independent. This assumption is often justified due to the partial compensation of  $r_{SBi}$  and  $\bar{Y}$ . Also, for practical applications of small-signal operation bias points at high injection are of little interest.

In a circular emitter the PDE can be reduced to a single variable equation which yields  $\underline{Z}_{Bi}$  in form of a hyper-geometric series [9]. As was shown in [35], restriction to the most important first-order frequency term leads to the compact representation

$$\underline{Z}_{Bi,lf} \cong \frac{1}{\frac{1}{R_{Bi,nc}^{circ}} + G + j\omega \frac{C_B}{3}} \approx R_{Bi,nc}^{circ} \left[ 1 - j\omega \frac{C_B R_{Bi,nc}^{circ}}{3} \right]. \quad (5.163)$$

The last expression results from  $1/R_{Bi,nc}^{circ} \gg G$ . Notice the factor  $g_w = 1/3$  instead of  $1/5$  in (5.50) for a long rectangular structure with contacts on two sides. Furthermore, for rectangular structures with surrounding contact  $g_w = 1/2.75$  was obtained from quasi-3D device simulation. Hence, for geometry scaling purposes in a compact model a simple interpolation is performed along the lines of (5.156),

$$g_\omega(b_E, l_E) = \frac{1}{5} + \left[ \frac{1-r}{2.75} + \frac{r}{3} - \frac{1}{5} \right] \frac{b_E}{l_E}, \quad (5.164)$$

with the normalized (corner rounding) radius  $r' = b_E r / (b_E \theta / 2)$ . This leads to the general geometry dependent formulation for the impedance

$$\underline{Z}_{Bi,lf} \approx R_{Bi,nc} [1 - j\omega g_\omega C_B R_{Bi,nc}] \quad (5.165)$$

with  $R_{Bi,nc}$  from (5.157). Figure 5.41 shows the geometry factor  $g_w$  for selected radii  $r'$ . The simple interpolation (5.164) is verified by device simulation of rectangular structures. The difference to rounded structures is quite small though.

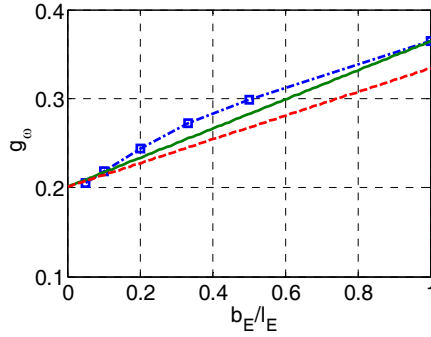


Fig. 5.41: Small-signal geometry factor  $g_\omega$  vs. aspect ratio: comparison between device simulation of rectangular structures (dotted line with symbols), (5.164) with  $r = 0$  for a square structure (solid line), and (5.164) with  $r = 1$  for a structure with corner rounding  $b_E/2$  (dashed line). The latter corresponds to a circular emitter for  $b_{E0}/l_{E0} = 1$ .

### C Large-signal transient operation

During a large-signal switching process in a 3D structure fundamentally the same physical mechanisms occur as explained in section 5.2.3. However, now the internal transistor is (dis)charged from both lateral directions, making it even more difficult to solve the underlying non-linear differential equation and to calculate a lumped effective resistance.

#### 5.4.3.2 Total base resistance for the 3D case

According to Fig. 5.39 the total base resistance consists of several components with possibly different bias and frequency dependence. For DC operation only  $R_{Bi}$  and, to some extent, the spacer portion  $R_{Bs}$  depend on bias. The other components may depend on the terminal variables only during transient operation. Generally, bipolar transistors offer a large variety of possible spatial arrangements of the base contact w.r.t. the emitter contact. Some examples are shown in Fig. 5.42. Be  $n_B$  and  $n_E$ , respectively, the number of base and emitter fingers, respectively. Then the DBC configuration in Fig. 5.42a is characterized by  $n_B = n_E + 1$ , the SBC configuration in Fig. 5.42b by  $n_B = n_E = 1$  or  $n_E = n_B + 1$ , and the FBC configuration in Fig. 5.42b by  $n_B = 1$  or  $n_B = 2$  regardless of  $n_E$ . These *contact*



*configurations* require for certain base resistance components (including  $R_{Bi}$ ) different descriptions of their layout dependence, which are discussed in the subsequent sections. In these considerations the internal transistor is defined by its effective dimensions ( $b_E, l_E$ ) according to section 5.3.1. Detailed descriptions and additional results can be found in [30, 31, 32, 33].

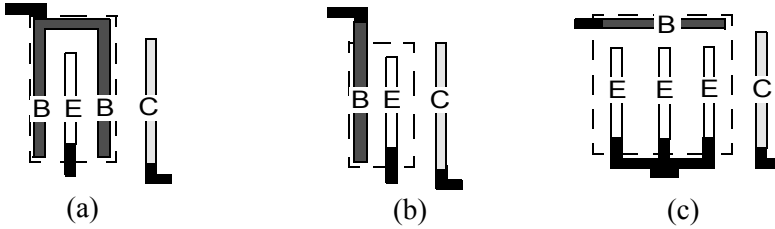


Fig. 5.42: Practically relevant BE contact configurations: (a) double base contact (DBC) structure with two base contacts per emitter contact finger in parallel; (b) single base contact (SBC) structure with one base contact and one emitter contact finger in parallel; (c) foreside base contact (FBC) structure with base contacts perpendicular to the emitter finger(s). The dashed box indicates the base poly layer.

The discussion and results below will be focused on the DC resistance. While for  $R_{Bi}$  the compact expression of the distributed resistance is the same (if current crowding is neglected) for both DC and AC operation this is generally not the case for  $R_{Bx}$ . This is due to the negligible DC current across the (reverse or low forward biased) external BC junction. Hence, there is no distributed current flow. However, under AC (or transient) operation the charging current of the various BC region related capacitance components can cause a non-negligible distributed current to flow at higher frequencies. Therefore, fundamentally  $R_{Bx}$  needs to be modelled similarly to (5.164) and (5.165) as a function of frequency (or time derivative). Note though that the total BC area contains regions with different values for the capacitance per area so that the distributed resistance of each region would need to be modelled separately in a compact model equivalent circuit.

### A Double base contact in parallel to the emitter

Figure 5.43a shows the magnified layout of a DBC structure for  $n_E = 1$  with the physical resistance components. The associated equivalent circuit follows from both quasi-3D device simulations and practical consider-

ations for constructing a compact model. Current flow and equipotential lines for DBC structures with different technology parameters have been presented in [30, 33]. Note that for the sake of generality the region ( $b_o + b_m$ ) in Fig. 5.39a has been partitioned into a silicided region ( $b_{sil}$ ) and a non-silicided poly region ( $b_p$ ) that includes both a poly-on-mono and poly-on-oxide portion.

For a DBC configuration  $b_E \leq l_E$  always holds, so that the boundary to the internal base can be considered as equipotential line (cf. [30,33]). Thus, the internal base resistance can be modelled by

$$R_{Bi} = r_{Sbi} \frac{b_E}{l_E} g_i(b_E, l_E) \psi(\eta) \quad (5.166)$$

with  $g_i$  from (5.156),  $\psi(\eta)$  from (5.151), and  $\eta$  from (5.158a) with  $g_h(b_E, l_E)$  from (5.158b).

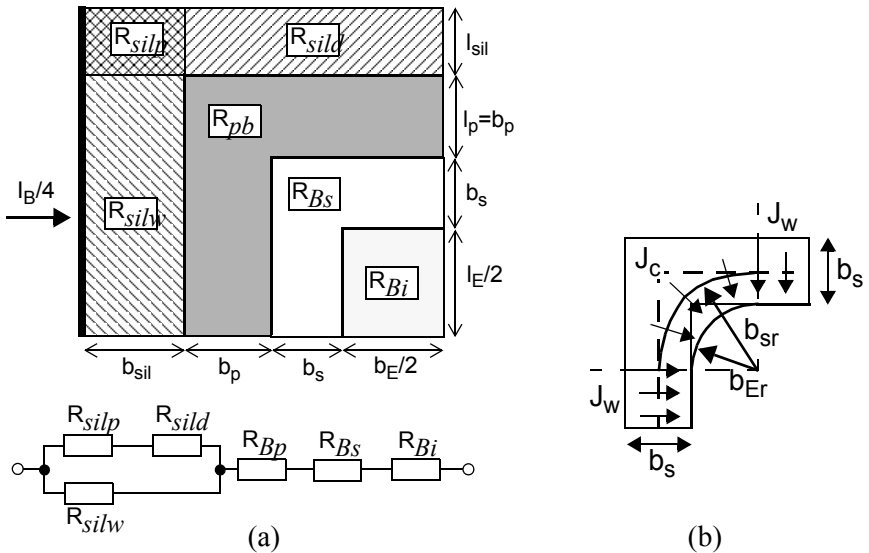


Fig. 5.43: (a) Schematic layout of a DBC structure with its physical resistance components and associated equivalent circuit. For symmetry reasons only a quarter of the structure is shown and used for device simulation. The assumed equipotential lines are indicated by thick lines. (b) Sketch of the spacer corner region layout with emitter window corner rounding (radius  $b_{Er}$ ) and relevant variables.

Except at the corners, the boundary between the spacer region and the poly-silicon or silicided region can also be considered as equipotential line due to the large difference between the sheet resistances of the spacer,  $r_{Ss}$ , and of the poly region,  $r_{Sp}$ . More detailed inspection of device simulation results (cf. Fig. 5.44) reveals that the equipotential lines at the corners look almost like quarter circles, similarly to realistic structures with an emitter window corner rounding radius  $b_{Er}$ . Therefore, the spacer resistance calculation is broken up into a “walled” component  $R_{sw}$  and a corner component  $R_{sc}$ .

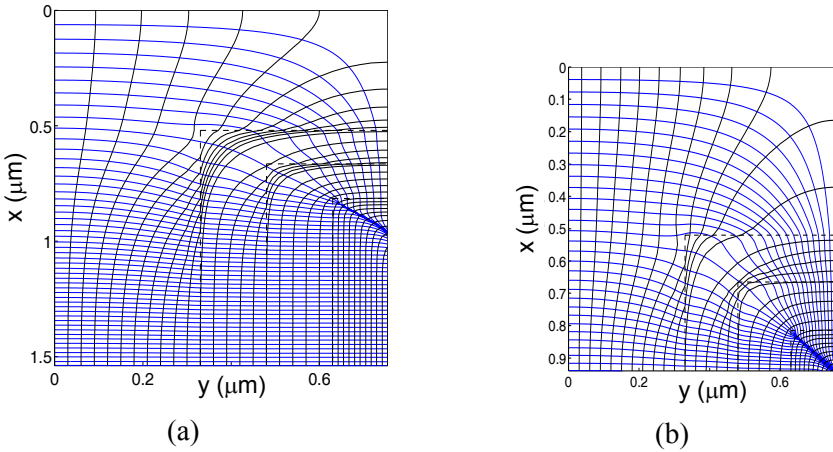


Fig. 5.44: Current flow and equipotential lines in a DBC structure for different emitter aspect ratios: (a)  $b_E/l_E = 0.25/1.45$  [33] and (b)  $b_E/l_E = 1$ .

The rounded equipotential lines suggest to model the corner resistance according to the theory for a circular structure (e.g. [37]):

$$R_{sc} = \frac{r_{Ss}}{2\pi} \ln\left(1 + \frac{b_s}{b_{Er}}\right) \quad \text{for } b_{Er} \leq \frac{b_E}{2}. \quad (5.167)$$

To avoid a discontinuity for  $b_{Er} \rightarrow 0$ , a sufficiently accurate approximation can also be obtained by using the average spacer corner radius  $b_{sr} = b_{Er} + b_{sr}/2$  (cf. Fig. 5.43), leading to

$$R_{sc} \cong \frac{r_{Ss}}{2\pi} \frac{b_s}{b_{sr}} \quad \text{for } b_{Er} \leq \frac{b_E}{2}. \quad (5.168)$$

The regular parallel current flow ( $J_w$  in Fig. 5.43b) across the front- and foreside is described by the walled component

$$R_{sw} = \frac{r_{Ss}}{2} \frac{b_s}{l_E + b_E - 4b_{Er}} \quad \text{for } l_E \geq b_E \geq 2b_{Er}. \quad (5.169)$$

The length in the denominator represents the width over which the parallel current flows. The factor 2 results from symmetry. If  $b_{Er} > b_E/2$  then  $b_{Er} = b_E/2$  is inserted. As the simulations show, the equipotential lines are even rounded for rectangular emitter windows with  $b_{Er} = 0$ . In this case,  $b_{sr}$  needs to be inserted directly into the *spacer related* equations and  $b_{Er} = b_{sr} - b_s/2$ . This case is only relevant for model verification from simulations or, e.g., for structures fabricated with electron beam lithography.

With the two components above, the total spacer resistance is given by

$$R_{Bs} = (1/R_{sc} + 1/R_{sw})^{-1}. \quad (5.170)$$

The sheet resistance  $r_{Sp}$  (20...200  $\Omega/\text{sq}$ ) of the non-silicided poly-silicon region adjacent to the spacer is usually much larger than the silicide sheet resistance  $r_{Ssil}$  (2...8  $\Omega/\text{sq}$ ) so that the current flows perpendicular through their boundary line. As a consequence, the same expressions as for the spacer resistance can be used if the proper variables are inserted:

$$R_{Bp} = \frac{r_{Sp}}{2\pi} \ln\left(1 + \frac{b_p}{b_{Er} + b_s}\right) \approx \frac{r_{Sp}}{2} \frac{b_p}{(l_E + b_E + 4b_s) + 2b_p}. \quad (5.171)$$

Since  $R_{Bp}$  is usually very small compared to  $R_{Bs}$  the simplified expression on the right can be used with its denominator corresponding to an average cross-sectional length.

The silicide resistance  $R_{Bsil}$  consists of two components. The first one represents the parallel (“walled”) current flow at the front side:

$$R_{silw} = \frac{r_{Ssil}}{2} \cdot \frac{b_{sil}}{l_E + 2(b_s + l_p)}. \quad (5.172)$$

The second component represents the current flowing to the emitter fore-side. It consists of a portion given by the access region,

$$R_{silp} = \frac{r_{Ssil} b_{sil}}{4 l_{sil}}, \quad (5.173)$$

and a distributed portion [30] with an average path length  $b_E/2 = b_s + b_p$ ,

$$R_{sild} = \frac{r_{Ssil}}{4} \sqrt{\frac{\delta}{1-\delta}} \coth \left[ \frac{b_E/2 + b_s + b_p}{l_{sil} \sqrt{\delta(1-\delta)}} \right]. \quad (5.174)$$

Due to the low value of  $r_{Ssil}$  the distributed current flow occurs entirely within the silicide. The value  $\delta = 0.15$  was confirmed in [32] by device simulation of a special test structure.

Finally, the contact resistance is given by

$$R_{Bc} = \frac{\rho_{Bc}}{2L_{Bc}} \quad (5.175)$$

with  $\rho_{Bc}$  as the length-specific contact resistivity. For the example shown in Fig. 5.43a the contact length is  $L_{Bc} = l_E + 2(b_s + l_p + l_{sil})$ .

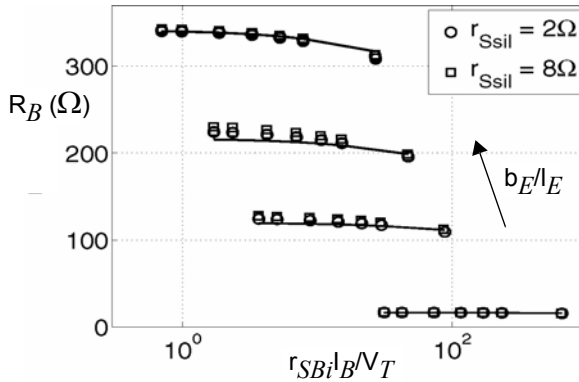


Fig. 5.45: Total base resistance of a DBC structure vs. normalized base current induced voltage drop for different emitter aspect ratios  $b_E/l_E = (0.02, 0.17, 0.38, 1)$ . Comparison between model (solid lines) and device simulation (symbols) for different silicide sheet resistances [33].

The total base resistance  $R_B$  is then calculated according to the equivalent circuit in Fig. 5.43a by inserting the expressions derived above. Figure 5.45 shows  $R_B$  for selected aspect ratios as a function of bias. The latter is caused by emitter current crowding in  $R_{Bi}$ . In all cases excellent agreement is obtained. In advanced HBTs the spacer and internal base resistance are usually much larger than the other components. For more results and discussions on both HBT and BJT structures see [30,33]. In

practice, a DBC structure is used mainly for speed, drive, and noise critical applications, such as amplifiers and drivers. In these cases, also multi-emitter transistors are used. The corresponding equations for the base resistance can be easily deduced from the ones presented above.

## **B Single base contact in parallel to the emitter**

With the introduction of the silicided base region it was tempting to eliminate the base contact between emitter and collector stripe, and only to maintain a silicided base layer with minimum width (cf. Figs. 5.39 and 5.42). This allows to reduce the external BC capacitance and collector resistance at the expense of an increased external base resistance. In how far the resulting single base contact is beneficial for the overall transistor performance depends on the widths and sheet resistances of the various base layers and the application. Figure 5.46 shows the enlarged layout of an SBC structure that is used for developing a compact base resistance model. The derivation principle follows the one used for DBC structures. Wherever adequate the already existing results are employed so that only the differences need to be discussed below in more detail.

Figure 5.47 provides a feeling for the current flow and equipotential lines especially on the opposite side (backside) of the base contact. For a square emitter (Fig. 5.47b), the flow line distribution hardly differs between front- and backside. However, at ratios  $b_E/l_E \ll 1$  the smaller number of flow lines at the backside can be clearly noticed in Fig. 5.47a. In this example ( $b_E/l_E = 1/10$ ) only about 38% of the total current supports the backside emitter region. This ratio decreases rapidly with smaller  $b_E/l_E$ . Thus, the equivalent circuit in 5.46 contains two separate resistance paths, one for the frontside (walled) current  $I_{front}$  and one for the distributed fore- and backside current  $I_{fore}$  and  $I_{back}$ .

Figure 5.47 provides a feeling for the current flow and equipotential lines especially on the opposite side (backside) of the base contact. For a square emitter (Fig. 5.47b), the flow line distribution hardly differs between front- and backside. However, at ratios  $b_E/l_E \ll 1$  the smaller number of flow lines at the backside can be clearly noticed in Fig. 5.47a. In this example ( $b_E/l_E = 1/10$ ) only about 38% of the total current supports the backside emitter region. This ratio decreases rapidly with smaller  $b_E/l_E$ . Thus, the equivalent circuit in 5.46 contains two separate resistance paths,

one for the frontside (walled) current  $I_{front}$  and one for the distributed fore- and backside current  $I_{fore}$  and  $I_{back}$ .

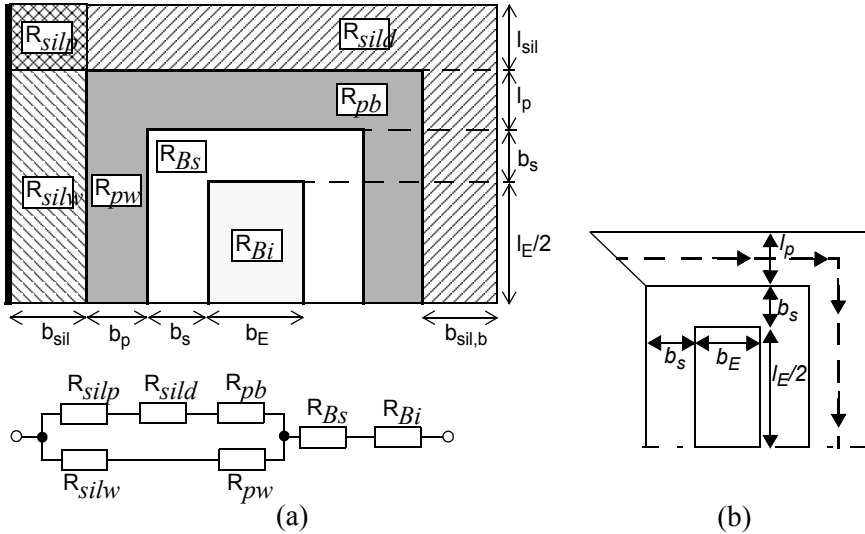


Fig. 5.46: (a) Schematic layout of an SBC structure with its physical resistance components and associated equivalent circuit. For symmetry reasons only half of the structure is shown that is also used for device simulation. The assumed equipotential lines are indicated by thick lines. (b) Sketch of the average path length for the current flowing through the poly region.

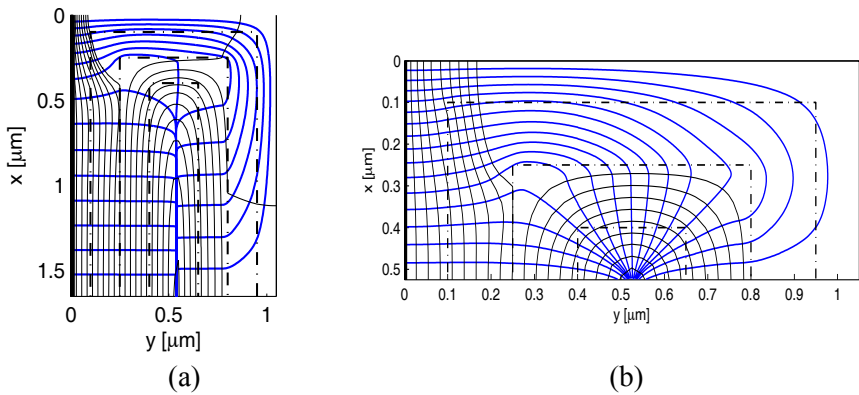


Fig. 5.47: Current flow and equipotential lines in an SBC structure for different emitter aspect ratios: (a)  $b_E/l_E = 0.1$  and (b)  $b_E/l_E = 1$ .

According to the equivalent circuit it is still assumed that the boundaries of the spacer region and the internal transistor are equipotential lines although this assumption obviously breaks down for (very) long emitter stripes; i.e. if  $I_{front}$  is significantly larger than  $I_{back}$ . Then, the internal base region is mostly supported from one side only, and the associated value for  $R_{Bi}$  increases by a factor of 4. This can be taken into account by modifying the corresponding geometry functions (5.156) and (5.158b) as follows:

$$g_i = \frac{1 + 3f_i}{12} - \left[ \frac{1-r}{28.45} + \frac{r}{8\pi} - \frac{1}{12} \right] \frac{b_E}{l_E} (1 - f_i), \quad (5.176)$$

$$g_\eta = \frac{17.84}{1 + 3f_i} - \left[ 11.9 \frac{b_E}{l_E} - 19.2 \left( \frac{b_E}{l_E} \right)^2 \right] (1 - f_i). \quad (5.176a)$$

Here,  $f_i$  is a suitable transition function, which is 0 for equally supported front- and backside but approaches 1 for a purely frontside support caused, e.g., by a large voltage drop in the surrounding silicide and poly region.

The main layout variables determining the transition between the above cases are the *resistance* of the silicide and poly region at the fore and front side in relation to the internal base resistance. Therefore, a transition variable  $u$  is defined that is given by the parallel circuit of the silicide and poly-Si resistance,

$$u = \ln \left( \frac{\left( \frac{b_{sil,b}}{r_{Ssil} (d_p + 2l_p + b_{sil,b})} + \frac{b_p}{r_{Sp} d_p} \right)^{-1}}{r_{Sbi} (b_E / l_E)} \right), \quad (5.177)$$

with  $d_p = 3b_s + b_E + l_p + l_E/2$  as average path length of the poly region (cf. Fig. 5.46b). The denominator containing relevant variables of the internal base resistance just serves as suitable normalization factor. For circuit optimization through sizing it is preferable to make  $f_i$  *smoothly* dependent on layout. This is accomplished by the following hyperbolic function

$$f_i(u) = a_{fi} [(u - u_0) + \sqrt{(u - u_0)^2 + e_{fi}}] - a_{fi} [(u - u_1) + \sqrt{(u - u_1)^2 + e_{fi}}] \quad (5.178)$$



with  $u_I = u_0 + 1/(2a_{fi})$  and the parameter vector  $(a_{fi}, e_{fi}, u_0)$ . The upper expression in  $f_i$  enables a smooth transition from the double-base case ( $f_i = 0$ ), while the lower expression smoothly limits the value of  $f_i$  to 1 for the single-base case.  $f_i$  was calculated from device simulation by subtracting  $R_{Bx}$  from  $R_B$ . The corresponding data points are shown in Fig. 5.48 and were used for fitting the parameters values of  $f_i$ . The resulting curve is also inserted in Fig. 5.48. Since very different sheet resistances and layouts have been included in the data, the fit parameters are believed to be valid for a wide layout and sheet resistance range that are relevant for practical applications.

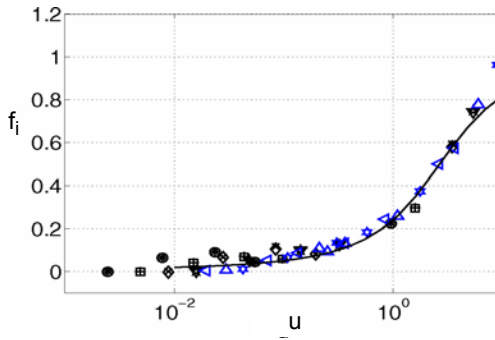


Fig. 5.48: Transition function  $f_i$  vs. transition variable  $u$  for all investigated SBC structures [33,31]. Data are indicated by symbols. The line represents a least-squares-fit with parameters  $a_{fi} = 0.65$ ,  $u_0 = 0.6$ ,  $e_{fi} = 2.5$ .

For the spacer resistance, the same considerations apply as before; i.e. the component on the back side disappears for support coming only from the frontside. Thus, (5.169) is modified by the transition function

$$R_{sw} = \frac{r_{Ss}}{2 - f_i} \frac{b_s}{l_E + b_E - 2b_{Er}} \quad \text{for } l_E \geq b_E \geq 2b_{Er}. \quad (5.179)$$

The corner components remains as in (5.168), and the total spacer resistance  $R_{Bs}$  is given by (5.170).

As long as  $r_{Ssil}$  is much smaller than  $r_{Sp}$  the current enters the poly-silicon region from the silicide region perpendicularly, even on the backside. Nevertheless, the poly resistance is split into a walled component

$$R_{pw} \cong r_{Sp} \frac{b_p}{l_E + 2b_s + b_p} \quad (5.180)$$

at the frontside, and a backside component

$$R_{pb} \cong r_{Sp} \frac{b_p}{l_E + 2b_E + 6b_s + 3b_p} \quad (5.181)$$

using average cross-section widths in the denominators (cf. Fig. 5.46b).

According to Fig. 5.46 the silicide resistance is split into a frontside (walled) component

$$R_{silw} = r_{Ssil} \frac{b_{sil}}{l_E + 2(b_s + l_p)}, \quad (5.182)$$

an access component for the corner, foreside and backside region,

$$R_{silp} = \frac{r_{Ssil} b_{sil}}{2 l_{sil}}, \quad (5.183)$$

and a distributed component

$$R_{sild} = \frac{r_{Ssil}}{2} \sqrt{\frac{\delta}{1-\delta}} \coth \left[ \frac{l_E/2 + b_E + 3(b_l + b_p)}{l_{sil} \sqrt{\delta(1-\delta)}} \right], \quad (5.184)$$

using an average current path length for the distributed current flow occurring in the silicide region only. If  $l_{sil} \neq b_{sil} \neq b_{silb}$ , their smallest value should be inserted.

The total base resistance  $R_B$  is again calculated according to the equivalent circuit in Fig. 5.46a by inserting the expressions given above. Figure 5.49 exhibits the bias dependence of  $R_B$  for selected aspect ratios and different silicide sheet resistance. In all cases good agreement is obtained with device simulation results. In practice, SBC structures are preferred for minimum size transistors and for reducing the BC (feedback) capacitance. For long stripes or (too) narrow silicide and poly layer width, their performance advantage over DBC structures needs to be carefully evaluated due to the possibly significant increase in base resistance.

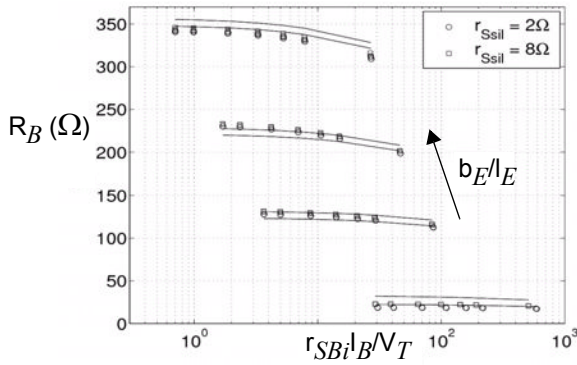


Fig. 5.49: Total base resistance of an SBC structure vs. normalized base current induced voltage drop for different emitter aspect ratios  $b_E/l_E = (0.02, 0.17, 0.38, 1)$ . Comparison between model (solid lines) and device simulation (symbols) [33].

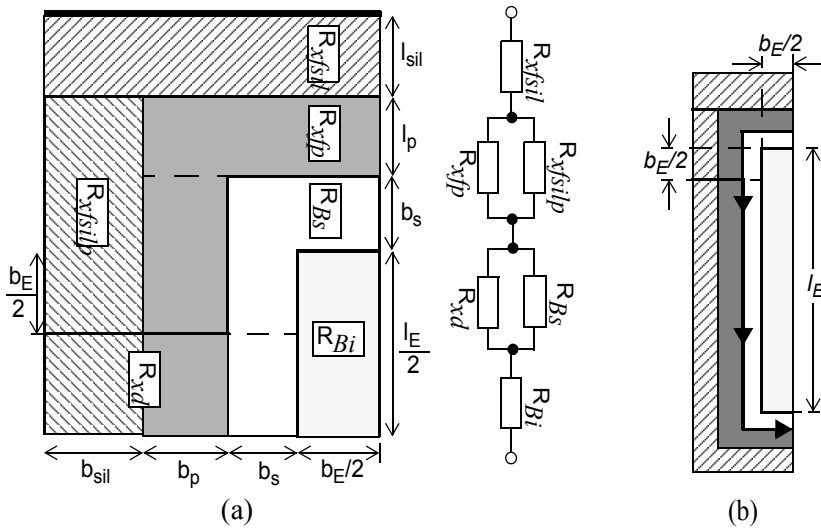


Fig. 5.50: (a) Schematic layout of an FBC structure with its physical resistance components and associated equivalent circuit. For symmetry reasons only a quarter of a structure with 2 base contacts is shown that is also used for device simulation. The assumed equipotential lines are indicated by thick lines. (b) Sketch of the spacer-to-poly boundary length for the distributed resistance calculation.

### C Base contacts perpendicular to the emitter

A reduction of the external collector resistance and partially also of BC capacitance can be achieved by arranging the base contact perpendicular to the emitter stripe. Figure 5.50 shows the enlarged layout of such an FBC structure that is used for developing a compact base resistance model. In principle, some of the equations for the SBC structure can be adapted to the case  $b_E \geq l_E$ . However, for longer devices the resistance is much more sensitive to errors in the description of the distributed region, and also the limiting case for  $R_{Bi}$  is different. The necessary modifications lead to a new set of equations that is described below. In this case also the number  $n_B$  of base contact stripes needs to be included in the expressions for the external base resistance.

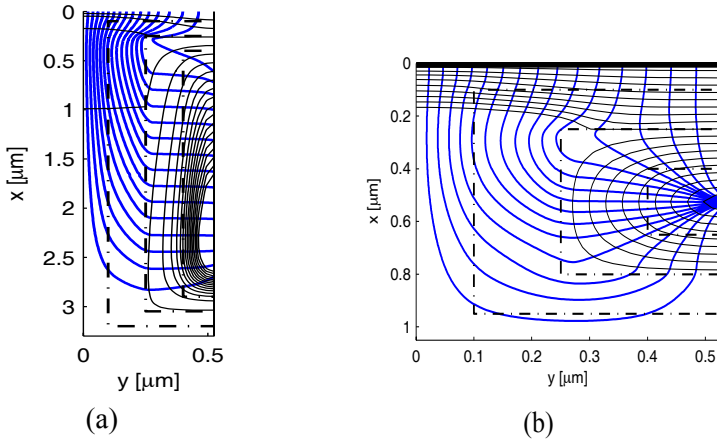


Fig. 5.51: Current flow and equipotential lines in a single contact FBC structure for different emitter aspect ratios: (a)  $b_E/l_E = 0.1$  and (b)  $b_E/l_E = 1$ .

The flow and equipotential lines in Fig. 5.51a clearly show the distributed character for longer stripes. The support of the frontside emitter region and, in case of  $n_B = 1$  also of the foreside opposite to the base contact, depends strongly on the sheet resistances and cross-sectional dimensions of the silicide and poly-silicon region. In order to enable calculations also for cases in which  $r_{Ssil}$  is not negligibly small compared to  $r_{Sp}$  the silicide and poly region are considered together with an effective sheet resistance

$$R_{Sx} = (b_{sil} + b_p) / \left( \frac{b_{sil}}{r_{Ssil}} + \frac{b_p}{r_{Sp}} \right). \quad (5.185)$$

Since the current flows perpendicularly across the spacer region boundary (cf. Fig. 5.51), the distributed length related resistance is calculated from the theory of contacts [38] as

$$R_{xd} = \frac{\eta_d}{n_B} \frac{1}{2} \sqrt{\frac{r_{Sx} r_{Ss} b_s}{b_{sil} + b_p}} \coth \left( \frac{l_d}{\delta_d l_{dT}} \right). \quad (5.186)$$

The factor 1/2 results from the symmetry of the structure along the emitter length direction;  $\eta_d (= 1)$  and  $\delta_d (= 1)$  are model parameters that can be adjusted for optimum accuracy; the length

$$l_d = \begin{cases} l_E + 2b_s & , n_B = 1 \\ l_E - b_E & , n_B = 2 \end{cases} \quad (5.187)$$

corresponds to the spacer-poly boundary length of the distributed region including the opposite corner region for  $n_B = 1$ . Finally, the transfer length

$$l_{dT} = \sqrt{\frac{r_{Ss}}{r_{Sx}} b_s (b_{sil} + b_p)} \quad (5.188)$$

determines the critical length

$$l_{d\infty} = \delta_d l_{dT} \quad (5.189)$$

beyond which  $R_{xd}$  is sufficiently close to its minimum value.

According to Fig. 5.50 and the associated flow and equipotential lines in Fig. 5.51, the emitter region at the corners and foresides is supported as in an SBC or DBC structure. The associated spacer resistance then consists of two components in parallel:

$$R_{Bs} = \left( \frac{f_{sc} \pi (b_{Er} + b_s / 2)}{r_{Ss} b_s} + \frac{f_{sc} 2(b_E - 2b_{Er})}{r_{Ss} b_s} \right)^{-1}. \quad (5.190)$$

The first portion represents the rounded corner region and the second the remaining “walled” and fore region. The dependence on  $n_B$  is described by the spacer corner function

$$f_{sc} = 2 + (n_B - 2) \left[ 1 - \left( \frac{l_{Ed}}{l_E} \right) \right]. \quad (5.191)$$

It takes into account that the current flowing across the opposite foreside and corner region is negligible for  $n_B = 1$  and sufficiently large  $l_E$  beyond a critical length

$$l_{E\infty} = \begin{cases} l_{d\infty} - 2b_s & , n_B = 1 \\ l_{d\infty} + b_E & , n_B = 2 \end{cases}. \quad (5.192)$$

Hence, the corresponding resistance components are eventually increased by a factor of 2. The smooth approach of the critical length by  $l_E$  is modelled through the variable

$$l_{Ed} = \frac{l_E}{[1 + (l_E/l_{E\infty})^{\beta_d}]^{1/\beta_d}}, \quad (5.193)$$

which is inserted in (5.191) and  $R_{Bi}$ . The factor  $\beta_d$  determines the transition from  $l_{Ed}$  to  $l_E$ , and has been determined from device simulations to  $\beta_d = 4$ .

The internal base resistance description is based on the existing equations (with  $f_i = 0$ ) as long as  $l_E$  is sufficiently short. Beyond a critical length, the reduced current support across the frontside can be simply taken into account by inserting  $l_{Ed}$  from (5.193) instead of  $l_E$ .

Finally, the access resistance for the foreside and corner region consists of the foreside related component for the silicided region (cf. Fig. 5.50),

$$R_{xfsil} = \frac{r_{Ssil}}{n_B} \frac{l_{sil}}{2(b_{sil} + b_p + b_s) + b_E}, \quad (5.194)$$

and of the component for the poly-on-mono region,

$$R_{xfp} = \frac{r_{Sp}}{n_B} \frac{l_p}{2(b_p + 2b_s + b_E)}, \quad (5.195)$$

which is in parallel to the component for the outer silicide portion

$$R_{xfsilp} = \frac{r_{Ssil}}{n_B} \frac{l_p + (b_s + b_E/2)/2}{2b_{sil}}. \quad (5.196)$$

The use of  $l_p + (b_s + b_E/2)/2$  in the last component is an approximation for the simultaneously occurring vertical and horizontal current flow.

The total base resistance  $R_B$  is then calculated according to the equivalent circuit in Fig. 5.50a by inserting the expressions given above. Figure 5.52 exhibits the bias dependence of  $R_B$  for selected aspect ratios. In all cases good agreement is obtained with device simulation results. In practice, FBC structures are preferred in transistors that require a reduction of the external BC capacitance for  $n_B = 1$  and, for short emitters, also of the external collector resistance. FBC structures are also often found for III-V HBTs. For long stripes or (too) narrow silicide and poly layer width, their performance is generally inferior to DBC or SBC structures with typical buried layer sheet resistance.

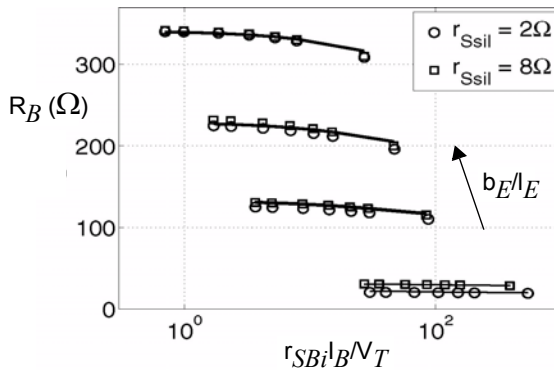


Fig. 5.52: Total base resistance of an FBC structure vs. normalized base current induced voltage drop for different emitter aspect ratios  $b_E/l_E = (0.02, 0.17, 0.38, 1.0)$ . Comparison between model (solid lines) and device simulation (symbols) [33].

#### 5.4.3.3 Summary

The derivations of the internal base impedance are based on the input circuit and, hence, describe the  $i_{Bi}(v_{B^*E})$  characteristic. The DC transfer current characteristic  $I_T(V_{B^*E})$  is reproduced correctly by the derived equations for  $R_{Bi}$  only if the DC current gain is bias independent. This assumption is made also in the literature although there sometimes the differential equation is formulated with the DC transfer current.

For AC operation the derived  $\underline{Z}_{Bi}$  reproduces the correct charging current, which in turn leads to a correct  $\underline{V}_{B'E'}$ , and, thus, the correct small-signal transfer current (as long as  $g_m$  is correct). For high-speed large-signal switching no closed-form solution can be obtained for both the internal input impedance and base resistance. Also, the typically used AC equivalent circuit *must not be used*. Therefore, in practical applications the DC base resistance is employed as first-order approximation for transient operation.

For DC operation the internal base resistance  $R_{Bi}$  has been described in a unified form, including rectangular and circular geometry. The same holds under AC operation for the internal base impedance  $\underline{Z}_{Bi}$  (within the limits of the theory). However, there are various options for constructing a lumped equivalent circuit to properly represent the exact solution. The same holds for high-speed transient operation.

#### 5.4.4 Avalanche pinch-in effect

So far, the vertical (1D) case and the lateral scaling of parameters have been considered for the avalanche effect in the collector. However, the avalanche current also influences the transistor terminal characteristics through special 2D and 3D effects. From a device point of view the behavior of the output characteristics as function of the transistor biasing mode is of high interest. Figure 5.53 exhibits qualitatively the results obtained for three characteristic cases. Beyond a certain voltage  $V_{CE}$  the collector current starts to increase significantly due to impact ionization in the collector. The point though at which the slope becomes infinite or the transistor enters the state of instability strongly depends on the biasing condition and, in particular, the flow of the holes created by impact ionization. Since an exhaustive discussion of the associated effects can be found in [39,29] only a summary of the most important device modeling related results will be given below along with qualitative explanations.



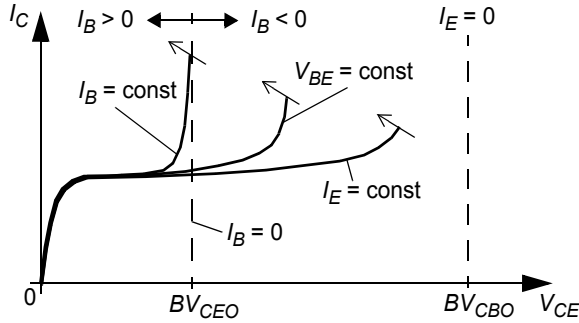


Fig. 5.53: Qualitative behavior of the output characteristics of a 3D transistor structure observed for the different biasing conditions  $I_B = \text{const}$ ,  $V_{BE} = \text{const}$ ,  $I_E = \text{const}$ . Note that  $BV_{CEO}$  and  $BV_{CBO}$  are defined at low injection.

Regardless of the biasing condition, holes created by impact ionization in the collector are accelerated towards the base. Once they enter the base region their flow direction is determined by the biasing condition. The most common case for device characterization is  $I_B = \text{const}$ . Since in this case the holes cannot leave the base through the terminal they are injected into the emitter. This additional base current  $(M-1)I_T$  in turn leads to an increase of the injected electron current  $I_T$  by  $B_f(M-1)I_T$  with  $B_f$  as forward current gain. Therefore, the collector current  $I_C = I_T + (M-1)I_T = MI_T$  can be expressed as a function of the base terminal current

$$I_C = I_B \frac{M}{1/B_f - (M-1)}. \quad (5.197)$$

For  $M-1 = 1/B_f$  the slope of  $I_C$  becomes infinite, and the transistor unstable. Accurate modeling of this point requires an accurate description of  $M$  and  $B_f$ . From a circuit application point of view though the condition  $I_B = \text{const}$  is of little importance compared to the conditions discussed next.

For  $V_{BE} = \text{const}$  the avalanche generated holes can also flow towards the base terminal (cf. Fig. 5.54). Increasing  $V_{CE}$  leads at some point to the situation that the avalanche current equals the back injection current resulting in  $I_B = 0$ . This is equivalent to the same condition,  $M-1 = 1/B_f$ , at which the instability for  $I_B = \text{const}$  occurs. However, while the corresponding open-

base breakdown voltage  $BV_{CEO}$  is the limit for  $I_B = \text{const}$ , it is not for  $V_{BE} = \text{const}$ .

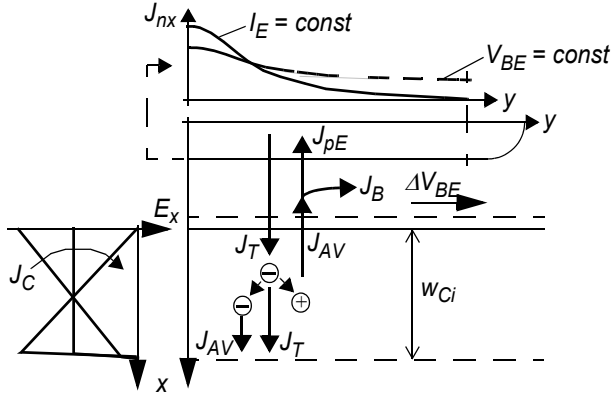


Fig. 5.54: Illustration of the collector avalanche effect in an npn transistor and its impact on the 2D hole current flow for  $V_{BE} = \text{const}$  or  $I_E = \text{const}$ . The most left figure shows the vertical electric field  $E_x$  for low to high current densities in the BC SCR which is assumed to be fully depleted ( $w_{BC} = w_{Ci}$ ). The upper figure contains the lateral distribution of the injected electron density  $J_{nx}$  for  $V_{BE} = \text{const}$  and  $I_E = \text{const}$ .

Further increase of  $V_{CE}$  causes  $I_B < 0$  and a voltage drop across the internal base towards the base terminal. In other words, the BE voltage across the emitter center junction now becomes larger than the terminal voltage. This is similar to emitter current crowding, but just in the opposite direction, and is called *pinch-in* effect. As a consequence, the (injected) transfer current density  $J_{nx}$  increases in the center. The corresponding increase of the backinjection current density reduces  $I_B$  and, hence, the lateral voltage drop somewhat. In addition, the electric field at the collector junction starts to decrease at sufficient current levels, leading to a decrease of the multiplication factor at the center. On the other hand, since now most of the holes come from the center region the average path becomes longer, which corresponds to an increased base resistance and voltage drop. If all generated holes would flow laterally, the resulting voltage drop would require an increase of backinjection; if all holes would flow into the emitter, however, the higher backinjection current requires a lateral voltage drop. Thus, a stable point, at an increased  $I_C$ , exists as long as the voltage drop

remains sufficiently small. A detailed analysis in [39, 29] shows that an instability occurs at the critical base current

$$I_{B, vb} \approx V_T / R_{bi}, \quad (5.198)$$

with  $R_{bi}$  as *small-signal* internal base resistance for the given emitter geometry. This corresponds to a voltage  $BV_{CE, vb} > BV_{CEO}$ .

For  $I_E = \text{const}$  and small ionization the avalanche generated holes can also flow towards the base terminal (cf. Fig. 5.54). Increasing  $V_{CE}$  leads again to the point  $I_B = 0$  and then to  $I_B < 0$ . The corresponding negative voltage drop also causes  $J_{nx}$  to increase at the center but now at the cost (i.e. reduction) of the current at the perimeter since the average emitter current must remain the same. Therefore, the increase of  $I_C$  and  $I_B$  at the same  $V_{CE}$  is lower than for  $V_{BE} = \text{const}$ . Compared to the latter case, the condition  $I_E = \text{const}$  leads also to a much larger ratio of center to perimeter current, i.e. to a stronger *pinch-in* effect. With increasing  $V_{CE}$  the center region is hogging more and more current. This process happens gradually, i.e. without positive feedback, until at  $I_B \rightarrow \infty$  the current density in the center becomes infinity while the current density outside the center line is zero. Such a 2D analysis obviously does not yield the actually observed infinite slope at a *finite* base current and CE voltage. Only a 3D analysis of the distributed current flow can explain this instability that is caused by a single *point* carrying the total current (*total pinch-in*). The corresponding critical base current can be approximated by [39]

$$I_{B, ie} \approx \frac{V_T}{R_{bi}} \left[ 1 + \left( \frac{b_E}{l_E} \right)^2 \right] \quad (5.199)$$

and occurs at a voltage  $BV_{CE, ie}$  beyond  $BV_{CE, vb}$ . Although the result for long stripes (i.e.  $b_E/l_E \ll 1$ ) is the same as for  $V_{BE} = \text{const}$ , the corresponding breakdown voltage  $BV_{CE, ie}$  is larger due to other, e.g., vertical (1D) effects. Note that total pinch-in does not necessarily happen at the center point but is determined by inhomogeneities in technological parameters under the emitter window that are caused by process tolerances.

In the discussion so far the influence of the electron current density on the field in the collector has been neglected. Since instability eventually leads to a high current density, the electric field at the BC junction decreases, leading to a reduction of the ionization rate. This acts as negative feed-

back effect until the field switches its slope and builds up again at the buried layer. The collapsing field in the center region also causes a lateral field that aids current spreading just beyond the BC junction, leading to a reduction of the current density in the center. Another issue to be considered is the instability in multi-emitter transistors which are often used for power amplifiers and driver circuits. Here again, caused by process tolerances, one emitter finger will eventually take over the total current. However, total pinch-in within this single finger will usually cause destruction.

The highly inhomogeneous lateral distribution of current density, field, base resistance, and also charge is very difficult to describe by a *compact* model; i.e. a model consisting of a *single* transistor element. Therefore, the impact of instabilities due to the collector avalanche effect on transistor and circuit characteristics can only be accurately predicted by a multi-transistor model representing the distributed effects.

As a final note it should be mentioned that the snapback, indicated by the arrows in Fig. 5.53, can only occur with a load resistance attached to the collector terminal. The details of the snapback behavior are determined by several factors including lateral and vertical effects in the transistor [39] as well as self-heating.

#### 5.4.5 Base-collector region related capacitances

As shown in the schematic cross-section in Fig. 5.55a the BC capacitance of a fabricated transistor consists of several components. First, the continuation of the SIC beyond the emitter window, and often even beyond the spacer region, leads to a contribution to the external depletion capacitance  $C_{jCx}$  with the largest capacitance per area. This is visualized in Fig. 5.55b where the area specific capacitance is drawn across the lateral direction in parallel to the BC junction. Once the SIC doping decreases to the background doping the associated specific depletion capacitance decreases significantly. A good approximation of the resulting capacitance is

$$C_{jCx} = (A_{sic} - A_E) \bar{C}_{jCi}(V_{B'C}) + A_{jcx} \bar{C}_{jCx}(V_{B*C}). \quad (5.200)$$

Here,  $A_{sic} = (b_{E0} + 2b_{sic})(l_{E0} + 2b_{sic})$  is the total SIC area and  $A_{jcx}$  is the area of the remaining BC junction with the cross sectional width  $(b_s + b_{pm} -$

$b_{sic}$ ) on each side of the emitter as well as at the foresides. The actual calculation of  $A_{jcx}$  includes also possible junction corner regions.

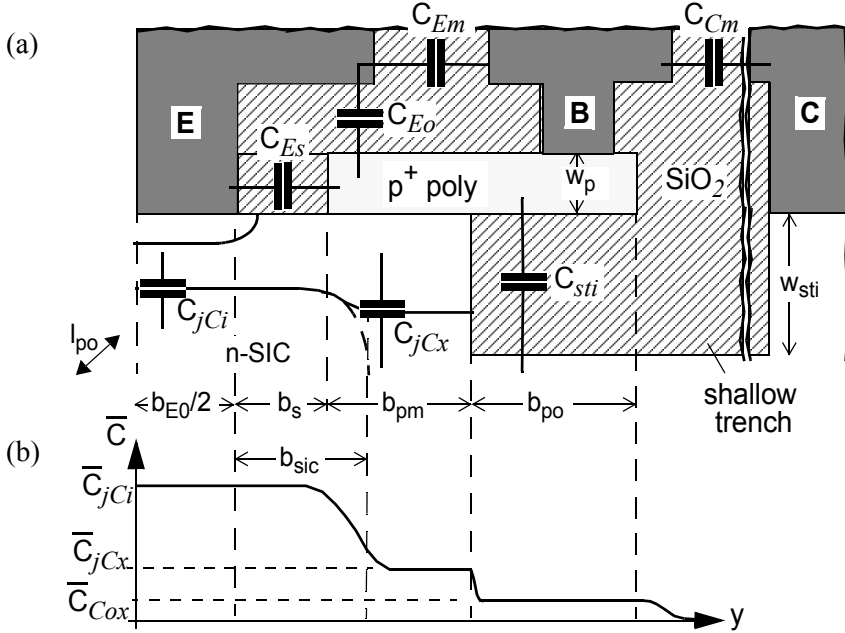


Fig. 5.55: (a) Schematic cross section of the base-collector and base-emitter region with relevant physical capacitance components and dimensions. (b) Schematic dependence of the vertically oriented base-collector capacitance per area along the lateral cross section direction.

In addition to the depletion capacitance, the shallow trench isolation (STI) results in the parasitic capacitance

$$C_{sti} = \frac{\epsilon_{SiO_2} \epsilon_0}{w_{sti}} A_{po} + C_{sti,f} \quad (5.201)$$

with  $\epsilon = \epsilon_{SiO_2} \epsilon_0$  and  $A_{po}$  as the total area of the plate-capacitor from the the poly-silicon on the STI. Furthermore,

$$C_{sti,f} = \epsilon \left[ 1.40 \left( \frac{w_p}{w_{sti}} \right)^{0.222} L_{po} + 4.12 \left( \frac{w_p}{w_{sti}} \right)^{0.728} w_p \right]. \quad (5.202)$$

is the fringing capacitance [40] The first term in the brackets represents the sidewall contribution with  $L_{po} = (l_{po} + 2b_{po})$  as the portion of the poly-base region perimeter length that does not face the emitter. The second term represents the corner contributions. (5.202) describes the contribution of each base contact between E and C contact. In technologies with deep trench isolation (DTI) one base contact is usually located over the trench so that its contribution to  $C_{sti,f}$  is negligible. Also, in single base transistors  $A_{po}$  and  $L_{po}$  need to be properly adapted to the actual layout.

Finally, in advanced technologies, the distance between the base and collector contact as well as the aspect ratio of contact (via) height and distance is causing a non-negligible parasitic metallization capacitance  $C_{Cm}$  between the contacts. An analytical calculation of this capacitance can be complicated due to possibly existing layers with different isolation materials (e.g. oxide and nitride) and the non-planar vertical structure.

#### 5.4.6 Base-emitter isolation capacitance

In addition to the depletion capacitances  $C_{jEi}$  and  $C_{jEp}$ , respectively, for the internal (bottom) and perimeter junction, respectively, the total BE capacitance contains parasitic elements the physical location of which is also displayed in Fig. 5.55a. Although an exact partitioning of the various components is difficult, their fundamental origin can be clearly distinguished.

The spacer related component  $C_{Es}$  is given by the spacer width and shape as well as by the isolation materials (e.g. oxide and nitride) employed. Due to the different shapes with possible undoped poly-fills, an accurate analytical description is quite difficult, and device simulation is preferred here. This capacitance is also very difficult to measure since it generally cannot be easily isolated from the perimeter depletion capacitance even if the voltage dependence of the latter is taken into account. Hence, in practice  $C_{Es}$  is very often included in  $C_{jEp}$  leading to a weaker *apparent* voltage dependence. The overlap between the emitter contact and the base poly-silicon causes the parasitic component  $C_{EO}$ . Based on the respective overlap width,  $C_{EO}$  can be roughly estimated from a plate-like and a fringing portion. Finally, the tight contact metal spacing in advanced processes gives rise to a metallization related component  $C_{Em}$  to which the same comments apply as before for  $C_{Cm}$ .

### 5.4.7 External collector resistance

Figure 5.56 contains a cross-section through the buried layer and substrate region along with the relevant components of the external collector resistance  $R_{Cx}$ , which is defined between the collector terminal contact C and the internal collector node C' (located at the interface between the internal (epitaxial) collector and the buried layer). Therefore  $R_{Cx}$  does *not* contain any contribution from the epitaxial layer under the emitter. Since analytical expressions for the various components of  $R_{Cx}$  depend on layout and mode of operation and, hence, can become quite complicated mostly a qualitative discussion will be given below with equations only for some selected cases. Complete working sets of equations have been implemented in TRADICA [41, 42] and used for many product designs over the years. The fundamental components can be discussed based on the most simple layout containing one collector and one emitter stripe (cf. Fig. 5.56).

The collector contact resistance is generally given by

$$R_{Cc} = \frac{\rho_{Cc}}{n_C A_{Cc}}, \quad (5.203)$$

with  $\rho_{Cc}$  as area specific resistance of the contact material stack,  $A_{Cc}$  as collector contact window area, and  $n_C$  as number of collector contact fingers. Often,  $\rho_{Cc}$  is about the same as the area specific emitter contact resistance. Directly in series to  $R_{Cc}$  is the sinker resistance  $R_{Cs}$  which is determined by the *vertical* doping profile under the contact (i.e. *not* the sinker sheet resistance!) and  $A_{Cc}$ . In general, the sinker resistance is quite difficult to measure separately from the contact component since the related regions are always fabricated together. Therefore, these two components are usually lumped together into a single component calculated by (5.203), but with an effective area specific “contact” resistance value.

Assuming a long structure (i.e.  $l_E \gg b_E$ ) allows to assume a homogeneous parallel current flow between the stripes. For DC operation, the resistance of the buried layer connecting the node C' with the sinker can then be calculated from the buried layer sheet resistance  $r_{Sbl}$ .

$$R_{Cb} = r_{Sbl} \frac{b_{EC}}{l_{EC}}, \quad (5.204)$$

with  $l_{EC}$  as (suitable) average length ( $= l_E$  for parallel flow lines).

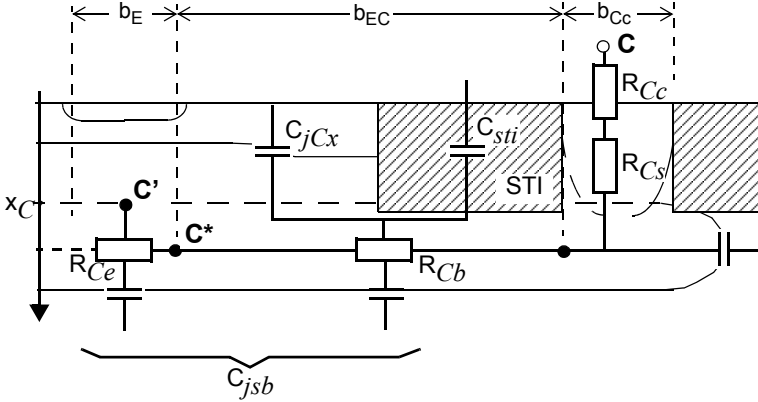


Fig. 5.56: The various components of the external collector resistance.

The resistance  $R_{Ce}$  under the emitter is determined by the distributed collector current flow there across the area at  $x_{C'}$  (cf. Fig. 5.56). Since for the most relevant case of forward active operation the transfer current source represents an output resistance that is much larger than the buried layer resistance, the impact of the distributed voltage drop on the current distribution is negligible. The current distribution itself at  $x_{C'}$  is given by possible current spreading at  $x_{C'}$ , which depends though on bias. Assuming a width  $b_{C'}$  and length  $l_{C'}$  of the distributed region under the emitter at  $x_{C'}$  as well as an average homogeneous current distribution allows for the 2D case to calculate the distributed resistance from a simple resistor chain. Here, the buried layer related component  $r_{Sbl}\Delta y/l_{C'}$  is much smaller than the value of the corresponding output resistance,  $r_{o,A}\Delta y/l_{C'}$ . Then, series expansion of the solution of the resulting differential equation yields

$$R_{Ce} = \frac{r_{Sbl} b_{C'}}{3 l_{C'}} \quad (5.205)$$

in analogy to the internal base resistance at low injection. Physically, as for the internal base resistance, the above resistor defines an effective voltage drop with its node  $C'$  located at a distance  $b_E/3$  away from the correspond-



ing location  $C^*$  (cf. Fig. 5.56). A rough approximation of the actual internal collector width is  $b_{C'} \approx b_E$  and  $I_{C'} \approx I_E$ .

Often, multi-emitter transistors are used for which the above calculation has to be extended. Figure 5.57 exhibits the schematic cross-section and relevant variables. In contrast to, e.g. the base current under the emitter, the collector current in the buried layer is distributed in a *discrete* rather than continuous way. A possible approach is to derive an equivalent lumped resistance from the voltage drop between the center of the emitter stripe next to the collector and the center of the innermost emitter:

$$\Delta V_d = R_d I_C = \Delta r I_C \left[ \left( 1 - \frac{1}{n_E} \right) + \left( 1 - \frac{2}{n_E} \right) + \dots + \left( 1 - \frac{n_E - 1}{n_E} \right) \right].$$

Here,  $\Delta r$  is the resistance between the center of two adjacent emitters and will be discussed later. Evaluating the sum in the brackets yields

$$R_d = \Delta r \frac{n_E - 1}{2}. \quad (5.206)$$

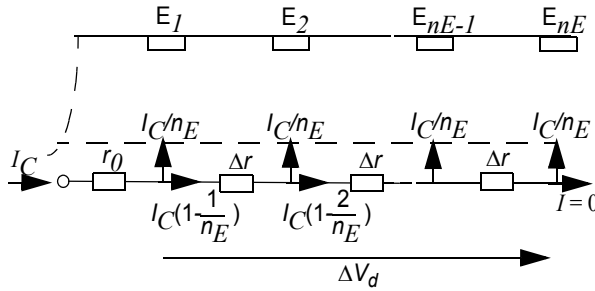


Fig. 5.57: Schematic representation of the discrete current distribution in a multi-emitter transistor with  $n_E$  emitter stripes and one parallel collector stripe. Note the difference to the continuous current distribution in the internal base region.  $r_0$  is the access resistance from the contact to the center of the first emitter stripe.

This equation is valid for one collector contact ( $n_C = 1$ ) on the side in parallel to  $n_E$  emitter stripes. The approach can also be applied directly to the case of  $n_C = 2$  with an even number of emitter stripes, if in (5.206) the variables  $I_C$  and  $n_E$  are replaced by  $I_C/2$  and  $n_E/2$ :

$$R_d = \Delta r \frac{n_E - 2}{8}. \quad (5.207)$$

Following the same lines, one can derive the resistance for  $n_C = 2$  and an odd number of emitter stripes:

$$R_d = \Delta r \frac{(n_E - 1)^2}{8n_E}. \quad (5.208)$$

Above equations contain several simplifications. First, the continuously distributed current flow under the emitter has been neglected. This can be taken into account in (5.206) to (5.208) by replacing the portion of  $\Delta r$  under the emitter by  $R_{Ce}$ . However,  $b_E$  is usually much smaller than the distance between stripes so that the impact of this modification is usually small, especially considering the simplification discussed next. Second, a parallel current flow is assumed. As visualized in Fig. 5.58a, the current flow is non-parallel in realistic structures similar to the base resistance discussed earlier. As a consequence, the equations for  $R_d$  given above need to be modified by a geometry function, leading to a smaller resistance. For other C and E contact arrangements, such as the one shown in Fig. 5.58b, an even more complicated current flow exists.

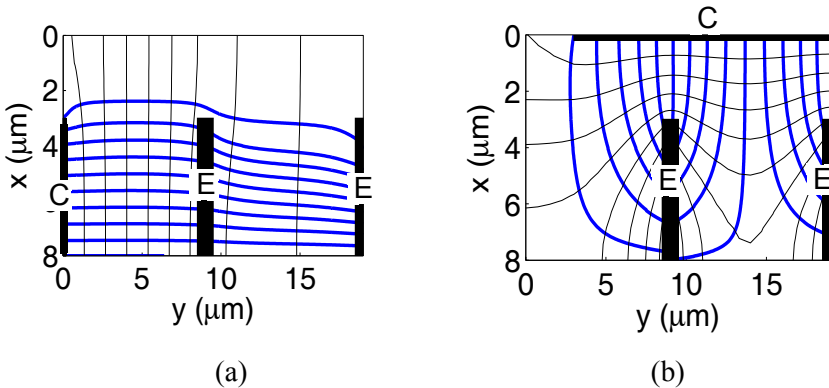


Fig. 5.58: Collector current flow lines in the buried layer of a transistor with three emitter stripes: (a) C contact parallel to the E stripes; (b) C contact perpendicular to the E stripes. In both cases,  $n_C = 2$  and only a quarter of the structure is displayed due to symmetry.

The non-parallel current flow can be taken into account semi-empirically based on device simulation results [41] or by Fourier-series based solutions [43]. The latter provide a more general and usually quite fast alternative but are less intuitive. A third simplification is to ignore DC current crowding. This is justified though because the voltage drop has to be very large in order to have any visible impact on the collector current. This is also true for the impact on capacitance values due to their weak square-root-like dependence on voltage.

The considerations so far have focused on the DC case. In contrast, during frequency dependent and transient operation the collector current flows distributed within the *entire* buried layer. This is caused by the charging current into the depletion and isolation capacitances of both the BC and CS region (cf. Fig. 5.56). While for DC operation  $R_{Cx}$  has relatively little impact in practical applications due to the series connection with the transfer current source, its influence can be significant for time and frequency dependent operation due to the associated time constants. If a DC resistance is used especially for the external buried layer regions represented by  $R_{Cbl}$  then the time constant associated with  $R_{Cbl}$  is incorrect by about a factor of 3. (The exact value depends on geometry; see discussion on base resistance). For sufficiently high frequencies or fast large-signal transients significant dynamic current crowding can occur, leading to an even larger error. Generally, an accurate description of  $R_{Cx}$  for AC and transient operation is of much higher importance than for DC operation. Thus, the distributed resistance value should be used for the buried layer components of  $R_{Cx}$  instead of their DC values. For very high frequencies, this resistance becomes also a function of frequency (cf. internal base resistance).

#### 5.4.8 Emitter resistance

Figure 5.59 displays the components of the emitter resistance  $R_E$  that are related to the contacting of the transistor; i.e. further contributions resulting from, e.g., metal lines that are required to connect the transistor to other devices are not considered within a transistor model. The metallization resistance  $R_{Em}$  includes the contributions from the silicide-poly and silicide-metal interfaces as well as from the silicide and metal bulk region. For a

slot contact  $R_{Em}$  can be scaled by the total contacting area  $A_{Em}$ . For via contacts, the resistance is determined mostly by the via (bulk) resistance and the via contact to the silicide due to the usually quite small via area.

The resistance  $R_{Ep}$  results from the highly doped poly-silicon bulk region. It scales with an area that is smaller than  $A_{Em}$  but somewhat larger than  $A_{E0}$ . Generally though  $R_{Ep}$  is small compared to the resistance of the interface between poly-silicon and mono-silicon,  $R_{Ei}$ . The latter scales with  $A_{E0}$  and is often the dominant component in  $R_E$ , especially for older (BJT) processes, due to a thin interfacial oxide.

Finally, the resistance  $R_{Eb}$  represents the mono-silicon bulk region. It is usually negligible for CED HBT processes due to their extremely shallow emitter junction and also for BJT processes. However, for mesa-type and LEC HBTs the resistance of their lightly doped emitter region close to the BE junction can become significant at higher bias when this region is no longer depleted.

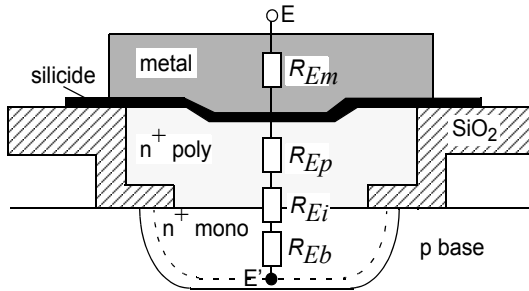


Fig. 5.59: Main components of the emitter resistance.

For many processes, measurements have shown a direct proportionality of  $R_E$  with the reciprocal emitter window area,

$$R_E = \rho_{Ec} / A_{E0}, \quad (5.209)$$

with  $\rho_{Ec}$  as effective area specific emitter contact resistivity. This proportionality indicates that  $R_E$  is mainly dominated by the poly-mono-silicon interface (and possibly also the bulk region), and that contributions from the other regions are of minor importance. However, this has to be verified for each process.

### 5.4.9 Collector-substrate capacitance

Figure 5.60 exhibits the two different structures used for isolation between devices. Modern high-speed processes offer a deep trench isolation (DTI) box with width  $b_{bl}$  and length  $l_{bl}$ . The trench is usually filled with poly-silicon and, hence, can have a floating potential. The latter can cause a channel along the DTI oxide and provide a conductive path between adjacent devices. Therefore, a channel-stopper is added at the bottom of the DTI. The trench depth is typically between 3 and 7  $\mu\text{m}$ . For a buried layer junction depth typically around 1.5 to 3  $\mu\text{m}$ , the space-charge region is generally completely contained within the DTI box.

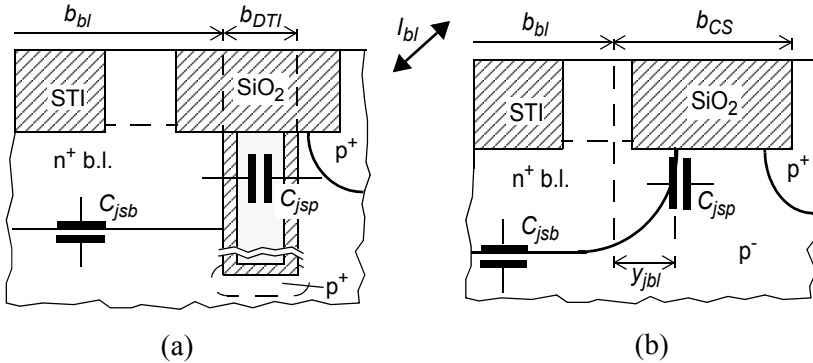


Fig. 5.60: Collector-substrate junction structure: (a) deep trench isolation and (b) junction isolation.  $b_{bl}$  and  $l_{bl}$  are the buried layer mask dimensions.

The junction isolation scheme exhibited in Fig. 5.60b is used in general in older process technologies, but is still also offered in advanced processes since a DTI can add significant fabrication cost. The larger footprint is due to (i) the lateral outdiffusion of the buried layer and (ii) the spacing required between the buried layer perimeter junction and the channel stopper. The spacing is optimized towards minimizing the perimeter CS junction capacitance, overall footprint, and CS breakdown current.

For the following discussion of the associated capacitance (and charge) a reverse biased CS junction is assumed. Also, the exact location and, in particular, the connection to the substrate contact at the surface is ignored for the time being. A forward biased junction and intra-device substrate coupling will be discussed separately later.

Both structures in Fig. 5.60 contain a bottom depletion capacitance component that is given by

$$C_{jSb} = A_{bl} \frac{C_{jS0,b}}{(1 - V_{SC}/V_{DSb})^{z_{sb}}} \quad (5.210)$$

with  $A_{bl} = b_{bl} l_{bl}$  as bottom area,  $C_{jS0,b}$  as the zero-bias area specific capacitance value of the bottom junction,  $V_{DSb}$  as built-in voltage,  $z_{sb}$  as exponent factor, and the controlling voltage  $V_{SC}$ . The connection to the internal collector node  $C'$  is justified if the sinker and contact resistance component dominate the value of  $r_{Cx}$ . Eq. (5.210) also implicitly assumes a lumped representation of  $C_{jSb}$  in an equivalent circuit. This can always be achieved for sufficiently small signals (below about 50 mV amplitude) if the corresponding series resistance description includes the distributed effect (cf. discussion in section 5.4.7).

The perimeter related substrate capacitance depends on the isolation scheme. For DTI, one can generally write

$$C_{jSp} = P_{bl} C_{CS,p}. \quad (5.211)$$

Here,  $C_{CS,p}$  is the perimeter related capacitance per length, the value of which is given by the material stack and layer thickness of the trench structure. Often,  $C_{CS,p}$  is assumed to be bias and temperature independent. The situation is somewhat more complicated for junction isolation. The component related to the perimeter depletion region can be written as

$$C_{jSp} = P_{SC} \frac{C_{jS0,p}}{(1 - V_{SC}/V_{DSp})^{z_{sp}}}. \quad (5.212)$$

Here,  $P_{SC} = P_{bl} + 4\gamma_{sc}$  is the average perimeter according to (5.121) with  $\gamma_{sc} = C_{jS0,p}/C_{jS0,b}$  as a reasonable approximation. A more accurate bias dependence would include punch-through to the channel-stopper region. In addition, a channel can exist along the shallow trench surface depending on the metallization and signal on top of the shallow trench or possible surface states at the oxide-substrate interface. Such a channel can extend over the whole distance  $b_{cs}$  or just a portion of it. A similar situation exists if the STI depth is smaller than the collector thickness  $w_C$ ; then the n doped epi region under the STI is directly connected to the buried layer potential. A

worst-case estimate for the associated capacitance is obtained by setting  $P_{SC} = P_{bl} + 4b_{sc}$  in (5.212).

#### 5.4.10 Parasitic substrate transistor

Any integrated bipolar transistor contains in general also a parasitic substrate transistor as shown in Fig. 5.61a for the example of an npn transistor. In this case, the npn-base layer acts as pnp-emitter, the npn-collector as pnp-base, and the npn-substrate as pnp-collector. In principle, also inverse operation is possible in which the npn-SC junction acts as emitter and the npn-base acts as collector. Depending on the process and the layout of the transistor the parasitic substrate transistor can turn on under certain electrical (circuit) conditions. This shall be discussed in more detail below for a parasitic pnp transistor with the help of device simulations using the doping profiles exhibited in Fig. 5.61b. During the discussion the voltage variables of the npn transistor will be used (unless otherwise specified) to characterize also the conditions of the pnp.

The most likely electrical condition for turning on the substrate transistor is a forward-biased BC junction, if the npn transistor is operated at very low *external* CE voltages. An example for this are power amplifiers, in which the transistor may be operated at low  $V_{CE}$  approaching  $V_{CEs}$ , which results in  $V_{B'C}/V_T \gg 1$ . Another condition for turning on the substrate transistor is a forward biased CS junction caused by voltage drops in the substrate (latch-up). However, the emitter efficiency and current gain are very low, making this operation much less likely. Quasi saturation, i.e. operation at high current densities, leads to a forward biased *internal* BC junction. However, the injection of holes into the buried layer is blocked by the large potential drop at the transition from the epitaxial collector to the buried layer that is caused by the still existing external reverse bias at the BC terminals. Fundamentally, one can distinguish between an “internal” substrate transistor  $T_{Si}$  with a *highly* doped buried layer in the base region (cf. dashed line in Fig. 5.61b), an “external” substrate transistor  $T_{Sx}$  with a *lightly* doped epi collector layer as base (cf. solid line in Fig. 5.61b), and a sinker substrate transistor  $T_{Ss}$  with a highly to moderately doped base region in lateral direction.





the terminal base current for a voltage controlled base or increase the internal base current for a current controlled base. (The effect on base current is similar to that of the avalanche current.)

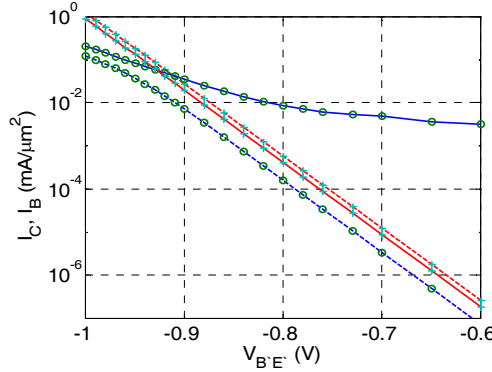


Fig. 5.62: Collector current (lines with o) and base current (lines with +) vs. pnp emitter-base voltage of the two different parasitic substrate transistor types: internal substrate transistor  $T_{Si}$  (dashed lines) and external substrate transistor  $T_{Sx}$  (solid lines).

Finally, the sinker transistor may turn on under similar conditions as for  $T_{Sx}$ . This occurs if the doping concentration at the bottom of the sinker becomes too small so that the sinker reaches its minimum lateral extension (i.e.  $T_{Ss}$  base width) there and the hole diffusion length reaches its maximum value. According to the discussion above, parasitic substrate transistor action can be avoided in a junction isolated process by a sufficiently high and deep collector sinker implant and by surrounding the npn by a sinker region. In a DTI process there is no (or extremely reduced) parasitic substrate transistor action since  $T_{Si}$  is unlikely to turn on.

For compact modeling, a possible substrate transistor needs to be taken into account. Based on the previous discussion, the model should represent  $T_{Sx}$ , which is the most likely component to turn on. The DC equivalent circuit consists of a transfer current source  $I_{TS}$  between S' and B as well as of two junction (diode) currents  $I_{jBC,S}$  and  $I_{jSC}$ , respectively, that represent the backinjection of electrons into the base and substrate, respectively. Since substrate transistor action is a second-order effect a simple transport formulation should be sufficient for describing the transfer current,

$$I_{TS} = I_{TSf} - I_{TSr} = I_{TSS} \left[ \exp\left(\frac{v_{BC}}{m_{Sf} V_T}\right) - \exp\left(\frac{v_{SC}}{m_{Sr} V_T}\right) \right], \quad (5.213)$$

which can be split into a forward and reverse component. The saturation current  $I_{TSS}$  as well as the ideality coefficients  $m_{Sf}$  and  $m_{Sr}$  are model parameters. The geometry scaling of  $I_{TSS}$  in  $T_{Sx}$  depends mainly on the distance  $b_x$  (cf. Fig. 5.61a). A SIC may increase this portion since the lateral field at the transition to the external collector region supports the outdiffusion of holes. Note, that the controlling voltage for forward operation is defined by the base terminal node since in advanced HBTs  $r_{Bx}$  is dominated by the spacer resistance, so that the node B\* is less suitable at least for  $T_{Sx}$ . Furthermore, the ideality coefficient can be used to include the influence of the transfer current of  $T_{Ss}$  and  $T_{Si}$  (if relevant).

The diode current  $I_{jBC,S}$  represents the backinjection of electrons from the pnp-base into the pnp-emitter region. It is different from  $I_{jBCx}$ , which represents the injection of holes in opposite direction (i.e. into the npn-collector) in a model without substrate transistor action. However, if  $T_{Sx}$  is included in a model hole injection is already being taken into account by  $I_{TSf}$ . Thus, in this case  $I_{jBCx}$  must be assigned the parameters of  $I_{jBC,S}$ . The diode current  $I_{jSC}$  always represents the injection of electrons into the lightly-doped substrate for a forward biased SC junction regardless of substrate *transistor action* or not.  $I_{jSC}$  can be described by the simple diode current law with an ideality coefficient and scales similarly to  $I_{TSS}$ .

The dynamic model for the substrate transistor consists of the BC and CS depletion charge as well as of a forward and inverse diffusion charge. The latter represent the minority charge storage in the lightly doped epitaxial collector, i.e. the pnp base region, for the most likely case of forward operation of  $T_{Sx}$ . The most simple description for the forward minority charge is obtained using a bias independent storage time. The geometry scaling of the minority charge follows the scaling of  $I_{TSS}$ .

#### 5.4.11 Intra-device substrate coupling

Since bulk silicon has a finite substrate resistivity  $\rho_{su}$ , any *integrated* device in silicon exhibits interactions with the substrate region in which it is physically embedded. So far, only the substrate junction itself was consid-

ered. However, as shown in Fig. 5.63 the substrate-sided edge of the CS SCR is connected to a substrate contact on the wafer surface through a resistive path, regardless of the isolation scheme. Therefore, at least a resistance needs to be added in series to the CS depletion capacitance. At sufficiently high frequencies also the (high) permittivity  $\epsilon_{su}$  of the silicon substrate comes into play. In a more detailed picture each infinitesimal element with length  $\Delta l$  and cross section  $\Delta A$  of the substrate current path can be represented by a parallel circuit consisting of a resistance  $\Delta R = \rho_{su} \Delta l / \Delta A$  and a capacitance  $\Delta C = \epsilon_{su} \Delta A / \Delta l$ . The resulting non-zero *impedance* between the substrate sided edge of the CS SCR and the substrate contact describes the effect of *intra-device* substrate coupling. The substrate time constant  $\tau_{su} = \epsilon_{su} \rho_{su}$  is an indication for the frequency limit beyond which the capacitive behavior of the substrate should be taken into account. For instance, the corner frequency  $1/(2\pi \tau_{su})$  in a  $10 \Omega\text{cm}$  substrate is about 15 GHz.

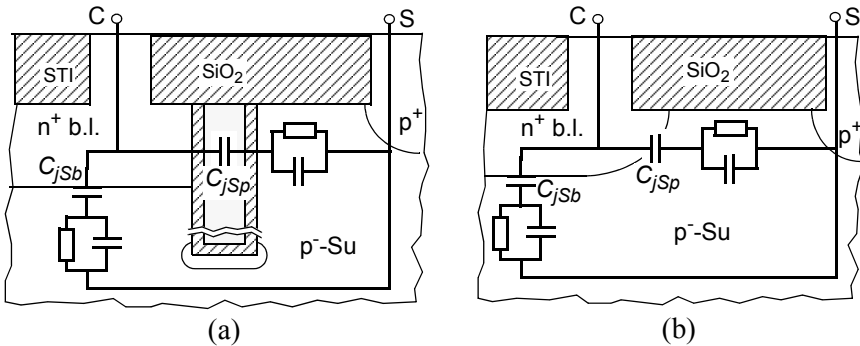


Fig. 5.63: Illustration of intra-device substrate coupling: (a) deep trench isolation and (b) junction isolation.

There have been numerous papers on substrate coupling in general but most of them focus on interactions between devices (*inter-device* coupling). Of those few papers dealing with *intra-device* coupling by far most use numerical solution methods for calculating the coupling impedance (e.g., [44-46]). These methods are of very limited use for compact model generation and design kits. However, the inherently three-dimensional problem makes it very difficult to develop fast and *compact* solutions for calculating the elements of a substrate coupling network as a function of

structural parameters such as dimensions and location of substrate contact. Only very few attempts are known that have attacked this problem [47, 48].

The most simple physics-based equivalent circuits for substrate coupling are shown in Fig. 5.63. They consist of a separate impedance for each current path with significantly different current flow. Both the junction- and trench-isolated structure in Fig. 5.63 contain an impedance branch for the buried layer bottom region and a branch for the buried layer perimeter region. For calculating the element values of these branches a certain subset of typical and widely used locations and shapes of substrate contact can be analyzed employing simple physics-based equations in combination with conformal mapping techniques [48]. A comparison of the obtained expressions for the geometry dependent substrate bulk resistance and capacitance with 2D device simulations, representing a buried layer with a length much larger than its width, showed excellent agreement within  $\pm 5\%$ , but revealed already significant deviations for the approach presented in [47]. The conformal mapping approach can be used to simplify (i.e. minimize the number of nodes) of the network automatically based on a specified frequency limit and error in order to keep the complexity of the compact model at a reasonable level.

So far the discussion has centred on a single substrate terminal with a contact located close to the device and at a known location. This is in fact the only situation for which it is possible to generate the element values of a substrate coupling network up front and to deliver them within the transistor model card. Often though circuit designers eliminate this device substrate contact in order to, e.g., save chip space or reduce line lengths, and instead place a contact somewhere on the chip. As a consequence, location, distance and shape of the substrate contact become unknown for model generation. In other words, the substrate network can only be determined *after* the circuit layout has been completed. Then the substrate network generation falls into the responsibility of a post-layout parasitic extraction tool, such as those for inter-device coupling. Thus, compact modeling of intra-device coupling only needs to take into account a limited number of substrate contact locations and shapes.

## 5.5 Non-standard-scaling

In recent advanced SiGe HBT process generations deviations from the standard scaling behavior (5.118) have increasingly been observed for certain characteristics such as transfer current, depletion capacitances, internal base sheet resistance, transit time, or avalanche current, although not simultaneously for all of them. This non-standard behavior can be caused by a variety of effects and fabrication conditions. Some of the causes are sketched in Fig. 5.64 in terms of pn junction contours and doping variations along the lateral direction under the emitter window.

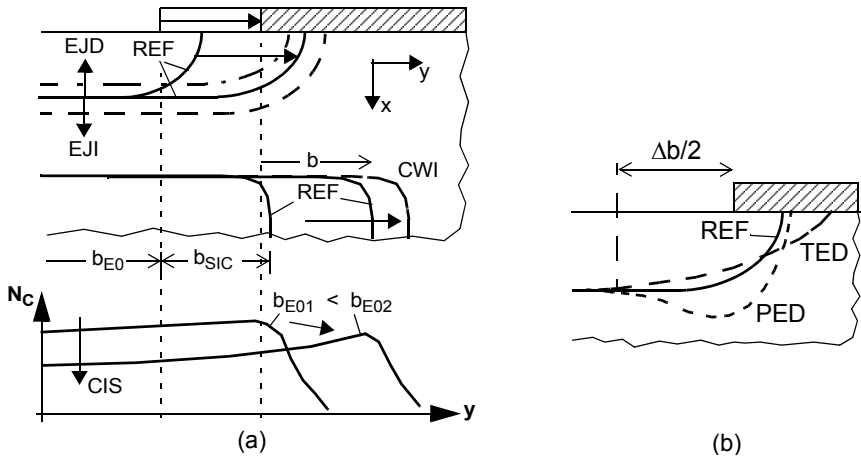


Fig. 5.64: Illustration of possible causes for non-standard scaling behavior. (a) pn junction contours (upper fig.) showing the variation of the planar BE junction (EJD, EJI) and the SIC extension with emitter width (CWI), and SIC doping change (lower fig.) under the emitter due to collector implant scattering (CIS). (b) junction contours showing the effect of transient enhanced diffusion (TED) and edge diffusion (PED) on emitter edge and perimeter junction region.

The junction contours of the reference structure with a profile exhibiting standard-scalable electrical behavior are indicated by REF. An emitter junction depth decrease (EJD) or increase (EJI) with a wider emitter or a pronounced emitter edge diffusion (PED) is sometimes observed [49]. This can happen under certain conditions for, e.g., emitter poly thickness and grain size, or the poly-mono interface etching. An extrinsic base implant too close to the emitter window can cause transient enhanced diffu-

sion (TED) [50,51]. Furthermore, the doping of the selectively implanted collector may vary with the window size (CWI) [52] due to, e.g., scattering of implanted doping atoms already at layers above the surface, which leads to a laterally non-equidistributed doping concentration (CIS) [53].

The effects impacting geometry scaling can be roughly divided into (i) purely layout dependent effects with layout *independent* process-specific parameters and (ii) vertical effects resulting from layout *dependent* specific parameters. The first case can still be covered by the standard-scaling law (5.118) through a suitable transformation between drawn and actual dimensions. However, the second case can lead to non-standard scaling behavior. Both cases may be further subdivided into emitter width and length dependent effects. A first systematic investigation attempting to separate and identify the various effects was performed in [54] for a production process [55] based on 2D device simulation that was calibrated to experimental data. The respective results and conclusions are summarized below. Note that (5.121) applies with  $P = P_0$  for  $l_{E0} \gg b_{E0}$  (2D simulation).

The scaling behavior of the collector current  $I_C$  at low injection is shown in Fig. 5.65. Changes in collector doping (CIS, CWI) turned out to have negligible influence here. PED results follow (5.121) but with a larger slope that indicates a higher perimeter injection due to the longer junction path there. For TED the decreasing BE junction depth towards the emitter edge leads to a slightly smaller perimeter injection (i.e. slope) at larger widths while the slope changes for  $b_{E0} < \Delta b$  (cf. Fig. 5.64b). However, for both PED and TED the area specific value  $I_{CA}$  remains the same as for REF, since the perimeter profile change disappears for large widths. The most pronounced deviation from standard scaling is observed for EJI. The encroachment of the high emitter doping into a region of significant base doping reduces the Gummel number for larger emitter widths, leading to an increase of  $I_C$  with width. The opposite case occurs for EJD, but the nonlinearity is much less pronounced. This is due to the (optimized) REF profile in which the emitter and base doping are already sufficiently separated (to reduce the BE depletion capacitance), leading to a smaller change in the Gummel number.

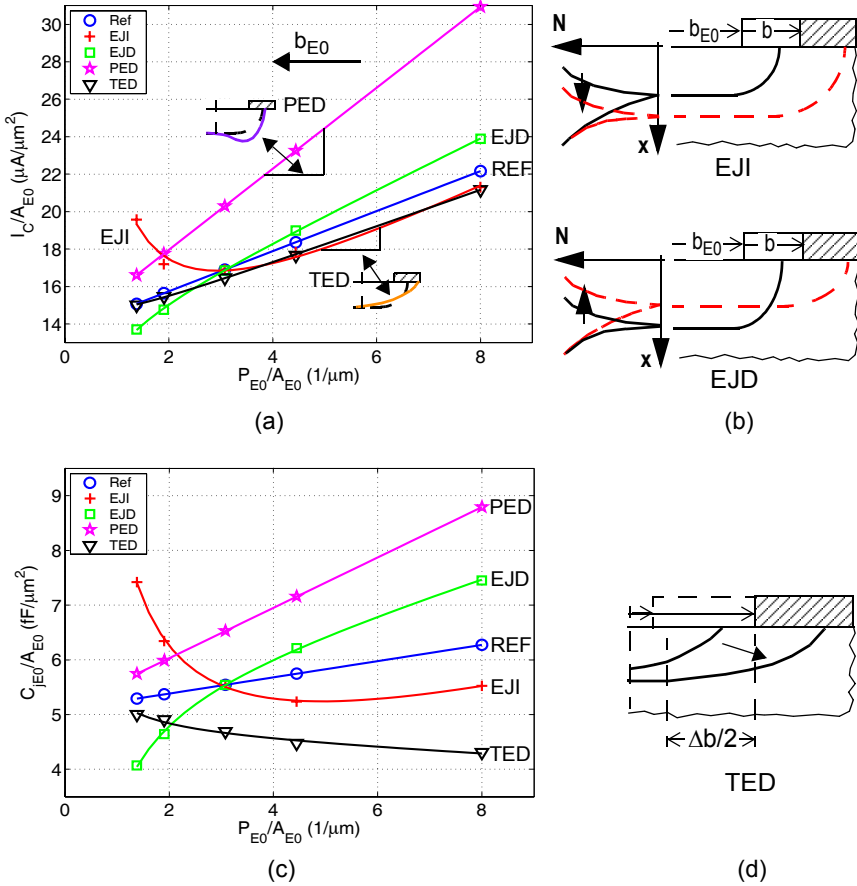


Fig. 5.65: Scaling behavior of (a) the collector current and (c) the zero-bias BE depletion capacitance. EJD, EJI and CWI are identical at reference width, and changes in the collector have no impact. Symbols represent device simulation results and lines the extended analytical scaling equation (5.214). For better illustration of the scaling behavior, figures (b) and (d) show a sketch of the doping contours and concentration for selected cases. Note that narrow widths correspond to large  $P_{E0}/A_{E0}$  ratios.

According to Fig. 5.65c the width dependent BE profile changes have a similar impact on the BE depletion capacitance as for  $I_C$ , but the non-standard behavior of PED, TED, and EJI is more pronounced since  $C_{jE}$  depends directly on (base) doping rather than on its integral (i.e. Gummel

number). The most noticeable observation is the negative slope for TED. Since in SiGe HBT processes the base doping towards the surface under the spacer becomes very small, the associated perimeter capacitance  $C_{jE,P}$  is also fairly small. For TED the junction contour recedes already within the emitter window (cf. Fig. 5.65d) leading to a decrease of the area specific term there compared to the value  $C_{jE,A}$  in the center of a wide emitter. Hence,  $C_{jE,A}A_{E0} > C_{jE}$  so that (5.121) only works with a negative value for  $C_{jE,P}$ , which corresponds to a negative slope. For sufficiently small widths  $b_{E0} < \Delta b/2$ , when the  $C_{jE,A}$  related portion has disappeared, the ratio of the (still constant) perimeter portion to the now width dependent edge portion increases, since the retracting junction depth within  $\Delta b/2$  is located at lower base doping concentrations. Thus, the absolute slope value decreases.

The observed non-standard scaling behavior needs to be described with sufficient accuracy by compact models in order to allow circuit optimization through device layout variation and, hence, full exploitation of a developed process technology for production. RF circuits in particular are very sensitive to the transistor layout used. As a consequence, “geometry scalable” compact models are necessary and have been provided in the past by several foundries using standard scaling (e.g., [41][56]). The inability though to describe non-standard scaling of certain parameters by (5.121) has increasingly caused foundries to disallow width dependent devices (and models) and to only provide length-scalable models. However, this can severely limit circuit optimization as well as predictive and statistical modeling and, thus, the overall return-on-investment for a process technology.

Although it often turns out that non-standard scaling effects could be eliminated or at least significantly be reduced by proper changes in the process flow, the problems are often either realized too late or considered to be too inconvenient or expensive to be fixed before process qualification. In either case, the burden of making the process cost-efficiently usable for circuit design is then put on compact modeling. In order to provide compact models for the same layout variations that designers have been used to with a standard-scalable process, parameters now would have to be extracted separately for many single-geometry transistors. This can become an extremely time consuming effort, which may still not lead to the same



layout flexibility and is likely to result in a loss of statistical modeling capability.

According to the observations in Fig. 5.65 and also in [54] for other variables, in some cases the geometry dependence still follows the standard scaling form (5.121) even though the doping profile varies with width. Therefore, those variables do not allow to electrically detect such profile changes. They only lead to a different set of specific parameters by using (5.121). However, looking at several variables simultaneously (e.g.,  $I_C$ ,  $C_{jE0}$ ,  $C_{jC0}$ ,  $r_{SBI0}$ ,  $BV_{CEO}$ , peak  $f_T$ ) allows to detect a profile change if at least one of those variables deviates from standard scaling. For instance, the depletion capacitances are often more sensitive to the corresponding doping changes than  $I_C$  or  $r_{SBI0}$  and may yield more conclusive information.

For doping profile changes that lead to a non-standard behavior of specific electrical parameters (5.120) can be easily extended:

$$\frac{Y_e}{A_0} = \frac{Y_{e,W}}{(P/A_0)} + Y_{e,A} + Y_{e,P} \frac{P}{A_0} + Y_{e,N} \left( \frac{P}{A_0} \right)^2. \quad (5.214)$$

Here,  $Y_{e,W}$  and  $Y_{e,N}$  are additional parameters representing a deviation from the standard scaling (i.e. the two middle terms) for wide (index “W”) and for narrow (index “N”) devices. For a given process, these additional parameters are constant, and usually only one of them has to be determined. In order to be able both to fit the parameters with sufficient accuracy and to distinguish a doping change related trend from variations due to measurement uncertainty, a sufficient number of structures with significantly different  $P/A_0$  ratios needs to be measured. The results obtained by applying (5.214) to  $I_C$  and  $C_{jE0}$  data of the devices with larger variations are shown as solid lines in Fig. 5.65. Excellent agreement is obtained in all cases using just the  $Y_{e,W}$  parameter. In practice, long structures with negligible corner contributions would be used to determine the parameters.

The drawback of the proposed extension (5.214) is that it has to be applied to every specific (electrical) parameter rather than to the variables related more directly to the origin of the non-standard scaling behavior. Therefore, a more efficient and physics-based approach would be to apply (5.214) only to those technology parameters (TPs), such as base or collector doping, that are also used for predictive and statistical modeling (cf.

[57]). This way, all specific parameters related to the particular TP would automatically become width dependent if the proper relations have been established as described in e.g. [57]. This approach requires though an *identification* of the origin of non-standard scaling and a proper description of the related TPs from the observed scaling behavior.

## 5.6 References

- [1] E. Kamke, "Differentialgleichungen - Lösungsmethoden und Lösungen", Teil I (Differential equations - solution methods and solutions, Part I), Akad. Verlagsgesellschaft, Leipzig 1967.
- [2] R.L. Pritchard "Two-Dimensional Current Flow in Junction Transistors at High Frequencies", Proc. IRE, Vol. 46, pp. 1152-1160, 1958.
- [3] H.N. Ghosh, "A Distributed Model of Junction Transistor and its Application in the Prediction of the Emitter-Base Diode Characteristic, Base Impedance, and Pulse Response of the Device", IEEE Trans. Electron Dev., Vol. ED-12, pp. 513-531, 1965.
- [4] J.R. Hauser, "The Effects of Distributed Base Potential on Emitter-Current Injection Density and Effective Base Resistance for Stripe Transistor Geometry", IEEE Trans. Electron Dev., Vol. ED-11, pp. 238-242, 1964.
- [5] H.-M. Rein und M. Schröter, "A Compact Physical Large-Signal Model for High-Speed Bipolar Transistors at High Current Densities - Part II: Two-Dimensional Model and Experimental Results", IEEE Trans. Electron Dev., Vol. ED-34, pp. 1752-1761, 1987.
- [6] H.-M. Rein and M. Schröter, "Experimental determination of the internal base sheet resistance of bipolar transistors under forward-bias conditions", Solid-State Electron., Vol. 34, pp. 301-308, 1991.
- [7] M. Versleijen, "Distributed High Frequency Effects in Bipolar Transistors", Proc. IEEE Bipolar Circuits and Technology Meeting, Minneapolis, 1991, S. 85-88.
- [8] T.-Y. Lee, "Model enhancement and parameter extraction for the MMSPICE/QBBJT model", Ph.D. thesis, U. of Florida, 1997.
- [9] G. Blasquez, J. Caminade, G. leGac, "Analysis of the effects of current crowding on noise of transistors with a circular geometry. Application to transistor with any given geometry", Physica 92B, pp. 313-329, 1977.

- [10] G. Blasquez, J. Caminade, K.M. van Vliet, "An accurate analysis of noise in rectangular bipolar transistors including current crowding", *Solid-State Electronics*, Vol. 23, pp. 423-431, 1980.
- [11] W. Liu, "Handbook of III-V heterojunction bipolar transistors", John Wiley & Sons, New York, 1998.
- [12] J.L. Moll, "Large-Signal Transient Response of Junction Transistors", *Proc. IRE*, Vol. 42, 1954, S. 1773-1784.
- [13] R. Beaufoy und J.J. Sparkes, "The Junction Transistor as a Charge-Controlled Device", *ATE J.* 13, 1957, S. 310-327.
- [14] C.-T. Sah, "Fundamentals of Solid-State Electronics", World Scientific, Singapore, 1991.
- [15] H.-M. Rein, "Edge injection and pinch-in effect of transistors with cylindrical geometry", *Electronics Lett.*, Vol. 4, No. 25, pp. 553-555, Dec. 1968.
- [16] H.-M. Rein, "The input resistance of transistors during fast switching", Dissertation (in German), University Stuttgart, 1968; s.auch: *Archiv der Elektrischen Übertragung (AEÜ)*, Vol. 22, pp. 322-336, 1968.
- [17] O. Manck, W.L. Engl, "Two-dimensional computer simulation for switching a bipolar transistor out of saturation", *IEEE Trans. Electron. Dev.* Vol. 22, pp. 339-347, 1975. See also:  
O. Manck, "Numerical analysis of switching behavior in a two-dimensional bipolar transistor", Dissertation (in German), RWTH Aachen, 1975.
- [18] H.-M. Rein, "Improving the Large Signal Models of Bipolar Transistors by Dividing the Intrinsic base into two Lateral Sections", *Electronics Lett.*, Vol. 13, pp. 40-41, 1977.
- [19] M. Schröter: "Physical models for high-speed silicon bipolar transistors - a comparing overview", Habilitation thesis (in German), Ruhr-University Bochum, 1994.
- [20] M. Schröter and H.-M. Rein, "Investigation of very fast and high-current transients in digital bipolar circuits by using a new compact model and a device simulator", *IEEE J. Solid-State Circuits*, Vol. 30, pp. 551-562, 1995.
- [21] A. Gover and A. Gaash, *Solid-State Electronics*, Vol. 19, p. 125, 1976.

- [22] H.-M. Rein, "A Simple Method for Separation of the Internal and External (Peripheral) Currents of Bipolar Transistors", *Solid-State Electronics*, Vol. 27, pp. 625-632, 1984.
- [23] M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors with special regard to high current densities and two-dimensional effects", Ph. D. thesis (in German), Ruhr-University Bochum, Bochum, Germany, 1988.
- [24] M. Schröter and D.J. Walkey, "Scaling considerations of bipolar transistors using 3D device simulation", *Proc. SISDEP*, pp. 210-213, Sept. 1995.
- [25] M. Schröter and D.J. Walkey, "Physical modeling of lateral scaling in bipolar transistors", *IEEE J. Solid-State Circuits*, Vol. 31, pp. 1484-1491, 1996 and Vol. 33, p. 171, 1998.
- [26] M. Schröter, H. Tran, "Two-/three-dimensional GICCR for Si/SiGe bipolar transistors" *Proc. WCM, International NanoTech Meeting, Anaheim (CA)*, pp. 99-104, 2005.
- [27] H. Tran, "Physics-based analytical modeling of SiGe heterojunction bipolar transistors for high-speed integrated circuits", Ph.D. thesis, University of Technology Dresden, Germany, 2006.
- [28] J. Cressler, "SiGe HBT reliability issues", ch. 4.11 in: "Silicon heterostructure Handbook", J. Cressler (ed.), CRC Press NY, 2005.
- [29] M. Rickelt and H.-M. Rein, "A novel transistor model for simulating avalanche-breakdown effects in Si bipolar circuits", *IEEE J. Solid-State Circuits*, Vol. 37, 2002, pp. 1184-1197.
- [30] M. Schröter, "Simulation and modeling of the low-frequency base resistance of bipolar transistors in dependence on current and geometry", *IEEE Trans. Electron Dev.*, Vol. 38, pp. 538-544, 1991.
- [31] M. Schröter, "Modeling of the low-frequency base resistance of single base contact bipolar transistors", *IEEE Trans. Electron Dev.*, Vol. 39, pp. 1966-1968, 1992.
- [32] S. Lehmann, M. Schröter, "Improved layout dependent modeling of the base resistance in advanced HBTs", *Proc. WCM, International NanoTech Meeting, Boston*, pp. 795-800, 2008.
- [33] M. Schröter, J. Krause, S. Lehmann, D. Celi, "Compact layout and bias dependent base resistance modeling for advanced SiGe HBTs", *IEEE Trans. Electron Devices*, Vol. 55, No. 7, pp. 1693-1701, 2008.

- [34] H.-M. Rein and M. Schröter, "Base spreading resistance of square emitter transistors and its dependence on current crowding", *IEEE Trans. Electron Dev.*, Vol. 36, pp. 770-773, 1989.
- [35] J. C. J. Paasschens, "Compact Modeling of the Noise of a Bipolar Transistor Under DC and AC Current Crowding Conditions", *IEEE Trans. on Electron Dev.*, Vol. 51, No. 9, 1483-1495, 2004.
- [36] J. Early, "Design theory of junction transistors", *Bell System Tech. J.*, Vol. 32, pp. 1271-1312, 1953.
- [37] A.B. Phillips, "Transistor Engineering", McGraw Hill, New York, 1962.
- [38] H. Murrmann and D. Widmann, "Current crowding on metal contacts to planar devices", *IEEE Trans. Electron. Devices*, Vol. ED-16, pp. 1022-1024, 1969.
- [39] M. Rickelt, "Modeling the breakdown behavior of Si/SiGe bipolar transistors in high-speed integrated circuits", Ph.D. thesis (in German), Ruhr-University Bochum, Germany, 2004.
- [40] T. Sakurai and K. Tamaru, "Simple formulas for two- and three-dimensional capacitances", *IEEE Trans. Electron Dev.*, Vol. 30, pp. 183-185, 1983.
- [41] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener and A. Koldehoff, "Physics- and process-based bipolar transistor modeling for integrated circuit design", *IEEE Journal of Solid-State Circuits*, Vol. 34, pp. 1136-1149, 1999.
- [42] Y. Zimmermann, K. Moebus, H. Wittkopf, M. Schröter, "TRADICA – an integrated modeling tool linking process and circuit design", *IEEE SCD*, 4 pages, 2009.
- [43] N. Kauffmann, C. Raya, F. Pourchon, S. Ortolland, D. Celi, "Determination of the collector resistance  $RC_x$  of bipolar transistors", *Proc. 5<sup>th</sup> HICUM Workshop*, Crolles, France, 2005. See also: C. Raya, N. Kauffmann, F. Pourchon, D. Celi, T. Zimmer, "Scalable approach for external collector resistance calculation", *Proc. Int'l. Conf. Microel. Test Stru. (ICMTS)*, Tokyo (Japan), pp. 101-106, 2007.
- [44] R. Gharpurey and R. Meyer, "Modeling and analysis of substrate coupling in integrated circuits", *IEEE J. Solid-State Circ.*, Vol. 31, pp. 344-353, 1996.

- [45] H.-M. Rein and M. Pfost, "Modeling and measurement of substrate coupling in Si-bipolar ICs up to 40 GHz", *IEEE J. Solid-State Circ.*, Vol. 33, pp. 582-591, 1998.
- [46] M. Pfost, "Modelling of substrate influence in high-speed silicon bipolar circuits", Ph.D. thesis (in German), Ruhr-University Bochum, Germany, 2000.
- [47] S. Fregonese, D. Celi, T. Zimmer, C. Maneux, P. Sulima, "A scalable substrate network for compact modeling of deep trench isolated HBT", *Solid-State Electronics*, Vol. 49, pp. 1623-1631, 2005.
- [48] K. Moebus, Y. Zimmermann and M. Schröter, "Schwarz-Christoffel Mapping for LDMOS and bipolar transistor modeling", *IEEE SCD*, 4 pages, 2009.
- [49] M. Kondo, H. Shimamoto, K. Washio, "Variation in emitter diffusion depth by TiSi<sub>2</sub> formation on poly-silicon emitters of Si bipolar transistors", *IEEE Trans. Electron Dev.* Vol. 48, No.9, pp. 2108-2116, 2001.
- [50] P. Stolk, H.-J. Grossmann, D. Eaglesham, D. Jacobson, C. Rafferty, G. Gilmer, M. Jaraiz, J. Poate, H. Luftmann, T. Haynes, "Physical mechanisms of transient-enhanced dopant diffusion in ion-implanted silicon", *J. Appl. Phys.*, Vol. 81, pp. 6031-6050, 1997.
- [51] M. Hashim, R. Lever, P. Ashburn, "2D simulation of the effect of transient-enhanced boron diffusion from base of silicon germanium HBT due to an extrinsic base implant", *Solid-State Electronics*, Vol. 43, pp. 131-140, 1999.
- [52] A. Stricker, B. Voegeli, N. Feilchenfeld, B. Rainey, M. Gautsch, "Improved vertical PNP collector-base breakdown using 2D Monte-Carlo TCAD simulations", *Proc. BCTM*, pp. 66-69, 2005.
- [53] N. Derrier, T. Schwartzmann, Internal report, ST Crolles, 2007.
- [54] M. Schröter, S. Lehmann, D. Celi, "Non-standard geometry scaling effects in high-frequency SiGe bipolar transistors", *Proc. WCM, International NanoTech Meeting*, Santa Clara, pp. 603-608, 2007.
- [55] H. Baudry, B. Szelag, F. Deléglise, M. Laurens, J. Mourier, F. Saguin, G. Troillard, A. Chantre, A. Monroy, "BiCMOS7RF: a highly-manufacturable 0.25 $\mu$ m BiCMOS RF-applications-dedicated technology using non selective SiGe:C epitaxy", *Proc. BCTM*, pp. 207-210, 2003.

- [56] K. Walter, B. Ebersman, D. Sunderland, G. Berg, G. Freeman, R. Groves, D. Jadus, D. Hareme, "A scaleable, statistical SPICE Gummel-Poon model for SiGe-HBTs", Proc. BCTM, pp. 32-35, 1997.
- [57] M Schröter, H. Wittkopf, W. Kraus, "Statistical modeling of bipolar transistors", Proc. Bipolar Circuits and Technology Meeting (BCTM), Santa Barbara (CA), pp. 54-61, 2005.

This page is intentionally left blank



## **Chapter 6**

# **Temperature Effects**

## 6.1 Overview

Electronic equipment must be operational over an ambient temperature range of  $-40$  to  $85$  °C for standard consumer applications,  $-40$  to  $125$  °C for automotive circuitry, and  $-55$  to  $125$  °C for military applications. For the latter two, the devices must even withstand junction temperatures up to  $175$  °C from device power dissipation or self-heating! In addition, potential cryogenic and space applications require electronic system functionality in a range from  $-196$  °C ( $= 77$  K) up to about  $100$  °C. While electronic design using a particular process technology, such as SiGe HBTs, is optimized for just one of the aforementioned temperature ranges, device models are often expected to be flexible enough to cover the entire range. However, at temperatures below about  $220$  K special physical effects come into play while on the other hand the most relevant applications in terms of volume and revenue operate above such temperatures. Moreover, due to self-heating effects (see later) most of today's devices and electronics in general work at increased temperatures. Therefore, it makes more sense to put the main focus on accurately modeling devices over the temperature range between  $-55$  to  $175$  °C. Special low-temperature effects can then be added on demand.

The discussion in the subsequent sections is organized as follows. First, the most relevant physical effects are briefly reviewed that are responsible for the temperature dependence of semiconductor devices. Then, the basic equations required for compact bipolar transistor modeling are derived based on the existing theory in the previous chapters. Finally, self-heating effects and their impact on device behavior will be discussed. In any case, a substantial simplification of the physical mechanisms is required to arrive at suitable expressions for compact models.

## 6.2 Physical effects

Temperature shifts in semiconductor device characteristics are determined by the temperature dependence of just a few basic physical parameters. The most important impact results from bandgap, mobility and saturation velocity while volume and interface or contact recombination only play a minor role (except possibly at high temperatures in space-craft

and remote planet lander surveyor applications). This also holds for freeze-out and trap assisted tunneling for temperatures above -55 °C. For a detailed discussion of physical effects in semiconductors and of the associated models the reader is referred to e.g., [1, 2].

There are various approximations in the literature (e.g. [3,4,5]) for analytically describing the experimental results of the bandgap energy  $W_g$  or voltage  $V_g = W_g/q$  as a function of temperature. The most widely-used approximation reads [3,4]

$$V_g(T) = V_{g,c}(0) - \frac{\beta_g T^2}{T + T_g} \quad (6.1)$$

The corresponding parameter values for silicon are given in Table 6.1. This classical formulation though has certain disadvantages when applied to current-voltage relations and also assumes a somewhat awkward expression when shifted from 0K to a more application relevant reference lattice temperature  $T_0$  (e.g. 300K). A more suitable form is

$$V_g(T) = V_{g,l}(T_0) - a_g(T - T_0), \quad (6.2)$$

which can be obtained by linearizing (6.1) at  $T_0$ . Detailed investigations of temperature sensitive circuits, such as bandgap references, have shown that a linear temperature dependence does not provide sufficient accuracy over the relevant temperature range. The non-linear formulation suggested in [5] has turned out to be both accurate and more suitable for compact models:

$$V_g(T) = V_g(0) + K_1 T \ln(T) + K_2 T. \quad (6.3)$$

Note, that  $T$  decreases faster towards zero than  $\ln(T)$  increases towards minus infinity, so that the equation assumes the finite value  $V_g(0)$  at  $T = 0$ . Above equation can be transformed to  $T_0$  resulting in the more convenient form

$$V_g(T) = V_g(T_0) + k_1 \frac{T}{T_0} \ln\left(\frac{T}{T_0}\right) + k_2 \left(\frac{T}{T_0} - 1\right) \quad (6.4)$$

with the parameters  $k_1 = K_1 T_0$ ,  $k_2 = K_2 T_0 + k_1 \ln(T_0)$ , and the bandgap voltage  $V_g(T_0) = V_g(0) + k_2$ . The corresponding parameter values listed in

Table 6.1 were calculated from an improved fit [6], which is more accurate over a larger temperature range than the original data set in [5].

The bandgap voltage obtained from the equations discussed above is compared in Fig. 6.1. A large temperature range was chosen to provide a better overview on the behavior of the equations. For instance, the lower boundary of 50K includes the liquid nitrogen case at 77K; the upper boundary includes automotive specifications at 175 °C.

eq. (6.1)	$V_{g,c}(0)$ [V]	$\beta_g$ [V/K]	$T_g$ [K]
	1.170	$4.73 \cdot 10^{-4}$	636
eq. (6.2)	$V_{g,l}(T_0)$ [V]	$a_g$ [V/K]	-
	1.1245	$2.546 \cdot 10^{-4}$	-
eq. (6.4)	$V_g(T_0)$ [V]	$k_1$ [V]	$k_2$ [V]
	1.1239	-0.025377	-0.053485

Table 6.1: Parameter values of different bandgap formulations for Silicon.

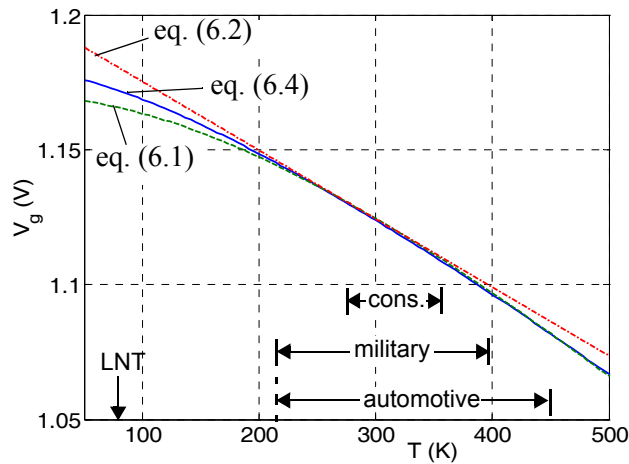


Fig. 6.1: Comparison of bandgap voltage approximations for Silicon (LNT = liquid nitrogen temperature).

The bandgap description influences the formulation of the effective intrinsic carrier density, which reads in general

$$n_{ie}^2(T) = n_{ie}^2(T_0) \left( \frac{T}{T_0} \right)^3 \exp \left[ \frac{V_{geff}(T_0)}{V_{T0}} - \frac{V_{geff}(T)}{V_T} \right]. \quad (6.5)$$

Here, the effective bandgap voltage,

$$V_{geff}(T) = V_g(T) - \Delta V_g(T), \quad (6.6)$$

includes the effect of bandgap narrowing (e.g. [7,2,8])

$$\Delta V_g(T) = \Delta V_g(T_0) + \delta_g \left( \frac{T}{T_0} - 1 \right). \quad (6.7)$$

The linear  $T$  dependence via  $\delta_g$  was found necessary from experimental current gain data of BJTs [9]. Inserting (6.6) along with (6.4) and (6.7) into (6.5) leads for the argument of the exponential function to

$$\frac{V_{geff}(T_0)}{V_{T0}} - \frac{V_{geff}(T)}{V_T} = \frac{V_{geff}(T_0) - k_2 - \delta_g \left( 1 - \frac{T_0}{T} \right) - \frac{k_1}{V_{T0}} \ln \left( \frac{T}{T_0} \right)}{V_{T0}}.$$

Noticing that  $V_{geff}(T_0) - k_2 - \delta_g = V_{geff}(0)$  finally yields

$$n_{ie}^2(T) = n_{ie}^2(T_0) \left( \frac{T}{T_0} \right)^{m_g} \exp \left[ \frac{V_{geff}(0)}{V_T} \left( \frac{T}{T_0} - 1 \right) \right] \quad (6.8)$$

with the constant

$$m_g = 3 - \frac{k_1}{V_{T0}} \quad (6.9)$$

and  $k_1 < 0$  from the bandgap voltage equation (6.4). Using the values in Table 6.1 gives  $m_g = 4.188$  for Si.

Besides bandgap, carrier transport plays an important role in the temperature dependence of device characteristics. According to [10], experimental results for the temperature dependence of the saturation drift velocity for  $T \geq 250$  K can be approximated by

$$v_{d,s}(T) = v_{d,s0} \left( \frac{T}{T_0} \right)^{-\zeta_{d,vs}} \quad (6.10)$$

with  $v_{d,s0} = v_{d,s}(T_0)$  and the coefficient  $\zeta_{d,vs}$  ( $= 0.87$  for electrons). Above relation is sometimes approximated by the linearized form

$$v_{d,s}(T) = v_{d,s0}[1 - \alpha_{d,vs}(T - T_0)],$$

with  $\alpha_{d,vs} = \zeta_{d,vs}/T_0$  as relative temperature coefficient given by a linearization at  $T_0$ .

The mobility of carriers in the semiconductor volume is determined by a variety of physical mechanisms (e.g. [10, 2]). For drift-diffusion simulation, these are usually combined into a single mobility expression via the Matthiessen rule. The resulting temperature dependent expression is by far too complicated for a compact model though. Focusing on the most important mechanisms of lattice and impurity scattering, the temperature dependence in the relevant range of about 200...600K can be approximated by the simple expression

$$\mu(T) = \mu(T_0) \left( \frac{T}{T_0} \right)^{-\zeta_{\mu T}}. \quad (6.11)$$

The exponent factor  $\zeta_{\mu T}$  is a function of the total doping concentration  $D$ . A rough approximation in silicon is [11, 9]

$$\zeta_{\mu T}(D) = \frac{2.5}{1 + (D/D_{\mu T})^{\beta_{\mu T}}}, \quad (6.12)$$

with the parameters  $D_{\mu T} = \{10^{17} \text{cm}^{-3} \text{ (n)}, 3.2 \cdot 10^{17} \text{cm}^{-3} \text{ (p)}\}$  and  $\beta_{\mu T} = 0.35$ . Note, that although these values have been adjusted to majority mobility they are also used for *minority* carriers. For a compact model,  $\zeta_{\mu T}$  needs to be adjusted as average value for each (neutral) device region. At low doping concentrations, the mobility is mostly determined by lattice scattering and decreases rapidly with temperature ( $\zeta_{\mu T} \approx 2...2.5$ ). Towards higher doping concentrations (charged) impurity scattering becomes increasingly important and reduces the temperature dependence, with  $\zeta_{\mu T} \rightarrow 0$  for very high concentrations like those in the emitter.

A very important physical effect is the generation of heat within the device mainly from carrier scattering. Hence, in regions with high current density and high gradients of the corresponding band edges such *self-heating* increases the temperature. This effect becomes increasingly important as on-chip integration density increases. Its impact on electrical device

characteristics as a consequence of *electro-thermal* interaction will be discussed in more detail in section 6.4.

According to the literature (e.g. [22]), recombination lifetimes in Shockley-Read-Hall and Auger processes depend, among others, on the thermal velocity and, thus, only weakly on temperature. The impact ionization rate determining the avalanche current decreases with temperature due to the increased contribution of random motion; i.e. a decreased influence of field directed motion. According to a more recent study for Si in [12], the parameters  $a_n$  and  $b_n$  in (3.120) vary slightly with temperature. Therefore, the simple approximations

$$a_n(T) = a_n(T_0)[1 + \zeta_{an}\Delta T], \quad (6.13a)$$

$$b_n(T) = b_n(T_0)[1 + \zeta_{bn}\Delta T] \quad (6.13b)$$

with  $\Delta T = T - T_0$  and the temperature coefficients  $\zeta_{an}$  and  $\zeta_{bn}$  seem justified and have worked well so far for BJTs and SiGe HBTs.

According to [13], the recombination velocity at poly- to mono-silicon interfaces is determined by the thermal velocity, tunneling and - for many grains - possibly also the diffusivity in the poly-silicon. Assuming an effective recombination velocity and high doping (i.e. relatively small coefficient  $\zeta_{\mu T}$  of the poly-silicon mobility) a rough approximation reads:

$$v_{pm}(T) \approx v_{pm0} \sqrt{T/T_0} \quad (6.14)$$

with  $v_{pm0} = v_{pm}(T_0)$ . A similar relation can also be assumed for the contact recombination velocity between silicon and metal.

The effects discussed so far are relevant for operation at temperatures beyond 200...250K. Towards low temperatures additional effects can gain influence. First of all, tunneling currents are generally negligible in well-designed SiGe HBTs. However, their temperature dependence is much weaker than that of junction currents so that the importance of tunneling increases towards low temperatures, leading to a decreasing operating range with ideal (or close to ideal) characteristics. In addition, hysteresis behavior has been observed in the output characteristics of advanced HBTs [14] that is caused by a negative differential conductance  $dI_B/dV_{BE}$ . Second, once carrier freeze-out occurs at low temperatures it reduces the free carrier concentration, resulting in an increase of series resistances and a decrease of  $Q_{p0}$ . In advanced SiGe HBTs the doping concentration in emit-

ter, base and buried layer are high enough to prevent freeze-out down to very low temperatures [15]. This is mainly due to the impurity level broadening and the screening of the carriers from the lattice atom potentials (a.k.a. high-doping effects). However, in power transistors with lightly-doped collectors freeze-out can already start there in the range of 200...250K, causing the internal collector resistance to increase. This in turn leads to a reduction in  $f_T$ . Finally, at low temperatures the contribution of (quasi-)ballistic transport increases [16].

The expressions presented above contain a lattice temperature  $T$  as variable that within a transistor structure is generally a function of position  $(x,y,z)$ . However, compact models use scalar variables that result from, e.g., integration or averaging over relevant device regions. As such single variable the junction temperature  $T_j$  is usually defined for the BE junction, since bipolar transistors are generally biased in forward mode and their characteristics are strongly determined by the temperature at the BE junction. The temperature distribution across the transistor structure will be discussed in more detail in section 6.4.

## 6.3 Temperature dependence of transistor characteristics

### 6.3.1 Charge storage

#### 6.3.1.1 Depletion charge

The key parameter for the temperature dependence of the depletion charges and capacitances is the diffusion (or built-in) voltage. Usually, the temperature dependence is derived from (3.154) by inserting (6.8) giving

$$V_D = V_D(T_0) - V_{geff}(0) \left( \frac{T}{T_0} - 1 \right) - m_g V_T \ln \left( \frac{T}{T_0} \right). \quad (6.15)$$

with  $V_{geff} = V_g - \Delta V_{g\Sigma}$ . The problem with this result is that for lower values of  $V_D(T_0)$  that correspond to lightly doped junctions  $V_D(T)$  can become negative within practical temperature ranges. This is because the *doping* concentrations are used instead of the majority carrier densities,



which is not correct anymore at high temperatures as was pointed out in [17]. Starting with (3.152) though a physics-based expression can be derived that is numerically stable for all temperatures.

At medium and high temperatures, the majority carrier density can be calculated from the neutrality condition at the edges of the SCR:

$$n_{n0} = p_{n0} + N_D, \quad p_{p0} = n_{p0} + N_A. \quad (6.16)$$

Eliminating the minority carrier densities by applying the  $pn$  product allows to determine the majority carrier densities directly

$$n_{n0} = \frac{N_D}{2} \left[ 1 + \sqrt{1 + \frac{4n_{ie}^2(x_n)}{N_D^2}} \right], \quad p_{p0} = \frac{N_A}{2} \left[ 1 + \sqrt{1 + \frac{4n_{ie}^2(-x_p)}{N_A^2}} \right]$$

with  $x_n$  and  $x_p$  as SCR edges. Inserting above expressions into (3.152) gives

$$V_D(T) = V_{Dj}(T) + V_T \ln \left( \frac{1}{4} \left[ 1 + \sqrt{1 + \frac{4n_{ie}^2(x_n)}{N_D^2}} \right] \left[ 1 + \sqrt{1 + \frac{4n_{ie}^2(-x_p)}{N_A^2}} \right] \right) \quad (6.17)$$

with the classical built-in voltage

$$V_{Dj} = V_T \ln \left( \frac{N_D N_A}{n_{i0}^2} \right) - \Delta V_{g\Sigma} = V_T \ln \left( \frac{N_D N_A}{\frac{2}{n_{ie}^2}} \right) \quad (6.18)$$

and

$$\overline{n_{ie}^2} = n_{ie}(x_n)n_{ie}(-x_p) = n_{i0}^2 \exp(\Delta V_{g\Sigma}/V_T). \quad (6.19)$$

Hence, for very high temperatures, the intrinsic carrier density term dominates the bracketed expressions in (6.17), so that  $V_D(T)$  physically correctly approaches zero. However, (6.17) is not yet suitable for compact modeling since the separate occurrence of  $n_{ie}$  at the SCR edges would require the knowledge of the respective bandgap reduction values. The goal is to find a formulation for  $V_D(T)$  that contains the corresponding model parameter  $V_D(T_0)$  and a "correction" term for a smooth extension to high temperatures.

The square root terms in (6.17) represent suitable smoothing functions between low and high temperature and give the classical result at medium temperatures. What is missing is an expression that is symmetrical in  $N_A$  and  $N_D$ , i.e. contains the product  $N_A N_D$ . A compromise regarding the transition behavior from medium to high temperatures is given by the approximation

$$\left[ 1 + \sqrt{1 + \frac{4n_{ie}^2(x_n)}{N_D^2}} \right] \left[ 1 + \sqrt{1 + \frac{4n_{ie}^2(-x_p)}{N_A^2}} \right] \approx \left( 1 + \sqrt{1 + 4 \frac{n_{i0}^2 \exp\left(\frac{\Delta V_{g\Sigma}}{V_T}\right)}{N_A N_D}} \right)^2,$$

which is exact for  $N_A = N_D$ . Recognizing that the last term under the square root equals  $\exp(-V_{Dj}/V_T)$  according to (6.18) and (6.19) permits to write (6.17) in a more consistent form required later:

$$V_D(T) = V_{Dj}(T) + 2V_T \ln \left( \frac{1}{2} \left[ 1 + \sqrt{1 + 4 \exp\left(-\frac{V_{Dj}(T)}{V_T}\right)} \right] \right). \quad (6.20)$$

If  $T$  approaches infinity then  $V_D$  correctly approaches zero. Note that this result does not include freeze-out and other effects occurring at very low temperatures. The next task now is to derive an expression for  $V_{Dj}(T)$  with  $V_{Dj}(T_0)$  as reference.

Inserting (6.5) into (6.18) and normalizing to  $\bar{n}_{ie}^2(T_0)$  gives

$$\exp\left(-\frac{V_{Dj}(T)}{V_T} + \frac{V_{Dj}(T_0)}{V_{T0}}\right) = \left(\frac{T}{T_0}\right)^3 \exp\left[-\frac{V_{geff}(T)}{V_T} + \frac{V_{geff}(T_0)}{V_{T0}}\right].$$

Moving the  $T^3$  term to the exponent allows to solve for the built-in voltage

$$-\frac{V_{Dj}(T)}{V_T} = -\frac{V_{Dj}(T_0)}{V_{T0}} + 3 \ln\left(\frac{T}{T_0}\right) - \frac{V_{geff}(T)}{V_T} + \frac{V_{geff}(T_0)}{V_{T0}}$$

and yields the *general transformation* for the classical built-in voltage equation with an *arbitrary* temperature dependent bandgap voltage:

$$V_{Dj}(T) = V_{Dj}(T_0) \frac{T}{T_0} - 3V_T \ln\left(\frac{T}{T_0}\right) + V_{geff}(T) - V_{geff}(T_0) \frac{T}{T_0} \quad (6.21)$$

In order to be able to practically use above expression and, hence, (6.20) a relation between  $V_{Dj}(T_0)$  and the actual model parameter  $V_D(T_0)$  has to

be established. This is done by solving (6.20) for  $V_{Dj}$  as function of  $V_D$ . Taking the exponent of (6.20) gives

$$\exp\left(\frac{V_D(T)}{V_T}\right) = \exp\left(\frac{V_{Dj}(T)}{V_T}\right) \frac{1}{4} \left[ 1 + \sqrt{1 + 4 \exp\left(-\frac{V_{Dj}(T)}{V_T}\right)} \right]^2,$$

which can be rearranged into a quadratic equation for  $\exp(V_{Dj}/V_T)$ . The solution reads

$$V_{Dj} = 2V_T \ln \left[ \exp\left(\frac{V_D}{2V_T}\right) - \exp\left(-\frac{V_D}{2V_T}\right) \right]. \quad (6.22)$$

Now, the procedure for calculating  $V_D(T)$  from  $V_D(T_0)$  can be specified. First, the value of the classical built-in voltage is calculated at nominal temperature  $T_0$  from (6.22),

$$V_{Dj}(T_0) = 2V_{T0} \ln \left[ \exp\left(\frac{V_D(T_0)}{2V_{T0}}\right) - \exp\left(-\frac{V_D(T_0)}{2V_{T0}}\right) \right]. \quad (6.23)$$

Next,  $V_{Dj}(T)$  is calculated using the above  $V_{Dj}(T_0)$  and the classical mapping (6.21) that contains the  $T$  dependent bandgap voltage (e.g. (6.4)). Finally, the desired  $V_D(T)$  is calculated from  $V_{Dj}(T)$  using (6.20). Note, that  $V_{geff}$  in (6.21) is different for every junction.

The zero-bias depletion capacitance can be expressed generally as  $C_{j0} \sim V_D^{-z}$  so that its temperature dependence can be directly calculated from that of  $V_D$ :

$$C_{j0}(T) = C_{j0}(T_0) \left( \frac{V_D(T_0)}{V_D(T)} \right)^z. \quad (6.24)$$

The temperature dependence of the depletion charge follows automatically by applying the above equations for  $V_D(T)$  and  $C_{j0}(T)$  and assuming that the exponent-factor  $z$  does not depend on temperature.

### 6.3.1.2 Mobile and minority charge

A temperature dependent description of the transit time can be based on the component equations derived in sections 3.5 and 4.4.2. However, in

some cases quite complicated equations would be obtained that are impractical for a compact model. For instance, it is generally difficult to determine the regional components with sufficient accuracy. Hence, a compact model will most likely have to be formulated in terms of lumped time constants. A compromise that has been found to be suitable for many different processes was proposed in [18] and implemented in HICUM. In more recent versions, a few improvements were made to that temperature formulation. Among those only the temperature dependence of the emitter storage time  $\tau_{Efo}$  requires a derivation here while all other changes follow straightforwardly from the physical models given in section 6.2.

Assuming in (4.147) the heterojunction to be aligned with the doping junction, the temperature dependence of  $\tau_{Efo}$  is given to first-order by the hole diffusivity in the neutral emitter, the contact recombination velocity and the current gain:

$$\tau_{Efo}(T) \approx \frac{1}{\beta_{f0}(T)} \left( \frac{w_E^2}{2\bar{\mu}_{pEm}(T)V_T} + \frac{w_E}{v_{pm}(T)} \right). \quad (6.25)$$

The temperature dependence of the SCR edge on the emitter side can be neglected for conventional doping profiles with highly doped emitters. For LEC transistors, proper device design is assumed in which the lower doped emitter region remains depleted under all practically relevant bias conditions. Inserting (6.11) gives for the diffusivity in the neutral emitter

$$\bar{\mu}_{pEm} V_T = \bar{\mu}_{pEm0} V_{T0} \left( \frac{T}{T_0} \right)^{\zeta_{tfE0}} \quad (6.26)$$

with  $\bar{\mu}_{pEm0} = \bar{\mu}_{pEm}(T_0)$  and  $\zeta_{tfE0} = 1 - \zeta_{\mu pE}$ . Inserting (6.14) for the recombination velocity yields

$$\tau_{Ef} = \frac{1}{\beta_{f0}} \left( \frac{w_E^2}{2\bar{\mu}_{pEm0} V_{T0} (T/T_0)^{\zeta_{tfE0}}} + \frac{w_E}{v_{pm0} (T/T_0)^{0.5}} \right).$$

For highly and moderately doped emitters  $\zeta_{mpE}$  is estimated to be between 0.24 ( $2 \cdot 10^{20} \text{ cm}^{-3}$ ) and 0.78 ( $2 \cdot 10^{18} \text{ cm}^{-3}$ ), respectively. Thus, assuming the value of  $\zeta_{tfE0}$  to be in the order of 0.5, the  $T/T_0$  ratios of the two components are similar. Furthermore, employing the temperature dependence of

the ideal DC current gain also for the ideal small-signal current gain, i.e.  $\beta_{f0}(T)/\beta_{f0}(T_0) \cong B_{fi0}(T)/B_{fi0}(T_0)$ , gives with (6.53c) and (6.6)

$$\tau_{Efo}(T) \cong \frac{\tau_{Efo}(T_0)}{\left(\frac{T}{T_0}\right)^{(\zeta_{ifE0} + \zeta_{Bfi})} \exp\left[\frac{\Delta V_{geff}(0)}{V_{T0}}\left(1 - \frac{T_0}{T}\right)\right]}. \quad (6.27)$$

The temperature dependence of the base-collector barrier can be derived as follows. Assuming that the change in material composition  $C$  extends over a sufficiently short spatial region, the bandgap variation across this region due to high-doping effects is negligible regardless of the barrier location  $x_{ch}$  (cf. sec. 4.1.3.3). Then, the valence band difference  $\Delta V_V$  between a point on the right (collector side, index “+”) and left (base side, index “-”) of  $x_{ch}$  is given by

$$\Delta V_V = (\Delta V_g(T, C) - \Delta \chi^-(T, C)) - (\Delta V_g^+(T, C) - \Delta \chi^+(T, C))$$

with  $\Delta V_g(T, C) = V_g(T, C) - V_{gr}$  and  $\Delta \chi(T, C) = \chi(T, C) - \chi_r$ , where  $V_{gr}$  and  $\chi_r$  are the bandgap voltage and electron affinity, respectively, in the reference bulk material (i.e. Si for a SiGe HBT). Describing the temperature and material dependent bandgap as

$$V_g(T, C) = V_g(T_0, C) f_g(T)$$

with  $f_g(T) = V_g(T)/V_g(T_0)$  as the normalized bandgap for bulk material (i.e.  $C = 0$ ), and the affinity by a linear temperature dependence,

$$\chi(T, C) = \chi(T_0, C) + a_\chi \left(\frac{T}{T_0} - 1\right),$$

leads to

$$\Delta V_V = \delta \Delta V_g(T_0, C) f_g(T) - \delta \Delta \chi(T, C)$$

Here  $\delta \Delta V_g$  and  $\delta \Delta \chi$  denote the difference between the left and right side of  $x_{ch}$ . Since  $C$  drops from a finite value on the left to zero on the right,  $\delta \Delta V_g$  and  $\delta \Delta \chi$  equal the changes of  $\Delta V_g$  and  $\Delta \chi$  with  $C$  on the left side of  $x_{ch}$ . Inserting the temperature dependence and combining terms then yields

$$\Delta V_V(T) = \Delta V_V(T_0) + a_{V2} \left( \frac{T}{T_0} - 1 \right) + a_{V1} \frac{T}{T_0} \ln \left( \frac{T}{T_0} \right) \quad (6.28)$$

with

$$a_{V2} = k_2 \frac{\Delta \tilde{V}_g(T_0, C)}{V_g(T_0)} - a_\chi, \quad a_{V1} = k_1 \frac{\Delta \tilde{V}_g(T_0, C)}{V_g(T_0)}. \quad (6.29)$$

This result can be applied directly to the barrier voltage parameter in (4.8) using the assumption made in sec. 4.1.3.3 that  $\Delta V_{Cb} = \Delta V_V/2$ .

### 6.3.1.3 Zero-bias hole charge

The general expression for the zero-bias hole charge reads

$$Q_{p0} = qA_E \int_0^{x_C} p_0(T) dx \quad (6.30)$$

with  $p_0$  as temperature dependent hole density at thermal equilibrium. The integration has generally to be performed over the entire (1D) transistor, i.e.  $x \in [0, x_C]$ . In practice though, only the base region will contribute to the integral, so that the limits can be replaced by the edges of the neutral base,  $x_{e0}$  and  $x_{c0}$ , at equilibrium.

Assuming complete ionization and abrupt SCR edges,  $p_0$  will then be equal to the doping concentration within  $x_{e0}$  and  $x_{c0}$  at any temperature, i.e.  $p_0 = N_B$ . However, a change of temperature will change the intrinsic carrier density and, thus, the electron density  $n$  at the SCR edges. For instance, an increase in  $T$  will also cause  $n$  to increase; this will lead to a smaller diffusion gradient across the junction and, hence, a smaller electric field to offset the diffusion in equilibrium. The consequence of a smaller field is obviously a decrease in the voltage drop across the junction and also a decrease of the SCR width. Since the voltage drop across the junction in equilibrium corresponds to the built-in voltage, which in turn determines the SCR width, the temperature dependence of  $Q_{p0}$  is determined by the temperature dependence of the built-in voltages. With these considerations, a simple derivation for the  $T$  dependence of  $Q_{p0}$  is given below assuming a spatially constant (average) base doping.

$Q_{p0}$  can be expressed through the  $T$  dependence of the SCR widths that are controlled by the built-in voltages:

$$Q_{p0}(T) = qA_EN_Bw_{b0}(T).$$

$w_{b0}$  is the neutral base width at thermal equilibrium and is given by

$$w_{b0}(T) = x_{c0}(T) - x_{e0}(T) = w_{b0}(T_0) - \Delta w_{be,b}(T) - \Delta w_{bc,b}(T)$$

with  $\Delta w_{be,b}$  and  $\Delta w_{bc,b}$  as the extension of the respective SCR into the base. The latter expression permits to write

$$Q_{p0}(T) = Q_{p0}(T_0) \left( 1 - \frac{\Delta w_{be,b}(T) + \Delta w_{bc,b}(T)}{w_{b0}(T_0)} \right) \quad (6.31)$$

with  $Q_{p0}(T_0) = qA_EN_Bw_{b0}(T_0)$  as reference variable. The  $T$  dependence of an SCR reads

$$w(T) = w(T_0) \left( \frac{V_D(T)}{V_D(T_0)} \right)^z,$$

and hence for the difference

$$\Delta w(T) = w(T) - w(T_0) = w(T_0) \left[ \left( \frac{V_D(T)}{V_D(T_0)} \right)^z - 1 \right]. \quad (6.32)$$

Inserting the respective expressions for the BE and BC SCR into (6.31) requires either the SCR widths (or respective ratios) to be defined as model parameters or certain assumptions to be made about their ratios. The first option is undesirable, since the widths are not measurable. The second option is to assume (i) that for properly designed profiles the zero-bias neutral base is still significantly determined by the BE SCR extension and (ii) similar voltage terms for the BE and BC SCR,

$$\left( \frac{V_{DEi}(T)}{V_{DEi}(T_0)} \right)^{z_{Ei}} \approx \left( \frac{V_{DCi}(T)}{V_{DCi}(T_0)} \right)^{z_{Ci}}.$$

Furthermore, defining the ratio

$$k = \frac{w_{bc,b0}}{w_{be,b0}} \cong \left( \frac{N_C}{N_B} \right)^{z_{Ci}} \left( \frac{N_E}{N_B} \right)^{z_{Ei}} \left( \frac{V_{DCi}(T_0)}{V_{DEi}(T_0)} \right)^{z_{Ei}},$$

and inserting above expressions via (6.32) into (6.31) give

$$\frac{Q_{p0}(T)}{Q_{p0}(T_0)} = 1 - (1 + k) \frac{w_{be, b0}(T_0)}{w_{b0}(T_0)} \left[ \left( \frac{V_{DEi}(T)}{V_{DEi}(T_0)} \right)^{z_{Ei}} - 1 \right]. \quad (6.33)$$

Finally, combining the prefactors of the built-in voltage into the parameter

$$k_{Qp0} = (1 + k) \frac{w_{be, b0}(T_0)}{w_{b0}(T_0)} \quad (6.34)$$

leads to

$$Q_{p0}(T) = Q_{p0}(T_0) \left[ 1 - k_{Qp0} \left[ \left( \frac{V_{DEi}(T)}{V_{DEi}(T_0)} \right)^{z_{Ei}} - 1 \right] \right], \quad (6.35)$$

As shown in Fig. 6.2, above expression yields a quite good approximation for both BJTs and HBTs if the factor  $k_{Qp0}$  is chosen properly according to the technology. For BJTs  $k_{Qp0} \approx 1$  while for HBTs  $k_{Qp0} \approx 0.25$  due to the high base doping concentration.

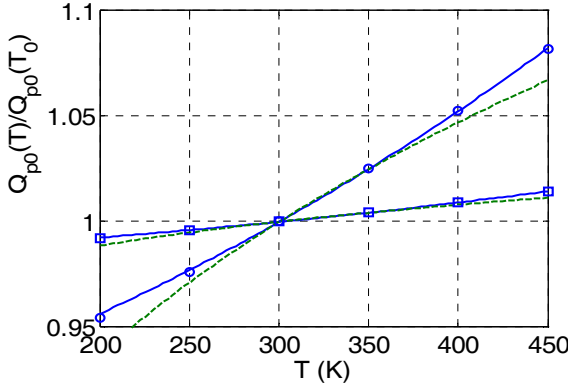


Fig. 6.2: Normalized temperature dependence of the zero-bias hole charge. Comparison between device simulation (symbols), expression (6.35) (solid lines) and expression (6.36) (dashed lines): (a) BJT (circles) with  $k_{Qp0} = 1.06$ ,  $\zeta_{Qp0} = 0.16$ ; (b) CED HBT (squares) with  $k_{Qp0} = 0.27$ ,  $\zeta_{Qp0} = 0.028$ .

A more convenient relation,

$$Q_{p0}(T) = Q_{p0}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{Qp0}}, \quad (6.36)$$

would be more compatible with the temperature dependent prefactor  $c_I$  of the transfer current and would make the associated parameter extraction



easier. However, as also shown in Fig. 6.2, this formulation is less accurate than (6.35).

## 6.3.2 Current components

### 6.3.2.1 Transfer current

According to (4.87) and (4.86) the transfer current is strongly temperature dependent via the intrinsic carrier density  $n_i$  and the exponential function of the controlling voltages. In addition, the weighted hole charge components all contribute differently to  $I_T(T)$ . First, the GICCR constant (4.86) is discussed since its dependence on the product of diffusivity and intrinsic carrier concentration is common to all descriptions of a transfer current. Inserting (6.5) and (6.11) yields

$$\frac{c_1(T)}{c_1(T_0)} = \left(\frac{T}{T_0}\right)^{-\zeta_{\mu nB}} \left(\frac{T}{T_0}\right) \left(\frac{T}{T_0}\right)^3 \exp \left[ \frac{V_{gBeff}(T_0)}{V_{T0}} - \frac{V_{gBeff}(T)}{V_T} \right] \quad (6.37)$$

with the *average effective* bandgap voltage in the (neutral) base,

$$V_{gBeff} = V_g - \Delta V_{gB}, \quad (6.38)$$

which contains a (possible) temperature dependence of the (average) bandgap-narrowing voltage  $\Delta V_{gB}$  in the neutral base region. After rearranging terms one obtains the general expression

$$\frac{c_1(T)}{c_1(T_0)} = \left(\frac{T}{T_0}\right)^{\zeta_{CT0}} \exp \left[ \frac{V_{gBeff}(T_0)}{V_{T0}} - \frac{V_{gBeff}(T)}{V_T} \right] \quad (6.39)$$

with the parameter

$$\zeta_{CT0} = m_g + 1 - \zeta_{\mu nB}, \quad (6.40)$$

the value of which depends on the selected bandgap voltage description via  $m_g$  ( $= 3$ ) at this point. Inserting (6.4) leads to

$$\frac{c_1(T)}{c_1(T_0)} = \left(\frac{T}{T_0}\right)^{\zeta_{CT0}} \exp \left[ \frac{V_{gBeff}(0)}{V_T} \left(\frac{T}{T_0} - 1\right) \right] \quad (6.41)$$

with  $m_g$  from (6.9). For modern Si-based technologies, typical values for  $\zeta_{\mu nB}$  are in the range of (0.2...1.5); inserting these together with the Si-value for  $m_g$  results in  $\zeta_{CT0} \approx 3.7...5$ . Similar considerations hold for a pnp transistor. Figure 6.3 shows  $c_I(T)$  with  $\zeta_{CT0}$  as parameter. Note that  $\zeta_{CT0} = 3.7$  corresponds to the classical linear bandgap temperature dependence and that the small difference observed in Fig. 6.3 is in fact relevant for designing bandgap reference circuits.

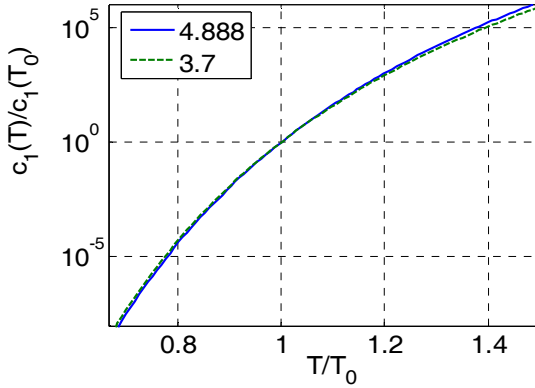


Fig. 6.3: Normalized variable  $c_I$  from (6.41) vs normalized temperature with fixed value of  $V_{gBeff} = 1.17V$  and  $\zeta_{CT0}$  as parameter.

Inserting  $c_I$  into the forward transfer current relation (4.87) yields

$$I_{Tf}(T) = \frac{c_I(T_0)}{Q_p(T)} \left( \frac{T}{T_0} \right)^{\zeta_{CT0}} \exp \left[ \frac{V_{gBeff}(0)}{V_T} \left( \frac{T}{T_0} - 1 \right) \right] \exp \left( \frac{V_{B'E}}{V_T} \right). \quad (6.42)$$

Accurate modeling of the temperature dependence requires measurement of the bandgap voltage. For low current densities,  $V_{B'C'} = 0$ , and a small  $V_{BE}$  range, the collector current as a function of temperature can be written as  $I_C(T) = I_{Tf}(T)$ . Extrapolation towards  $V_{B'E} = 0$  allows to define the temperature dependent saturation current

$$I_S(T) = I_S(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{CT}} \exp \left[ \frac{V_{gBeff}(0)}{V_T} \left( \frac{T}{T_0} - 1 \right) \right] \quad (6.43)$$

with

$$I_S(T_0) = \frac{c_{10}(T_0)}{Q_{p0}(T_0)}, \quad (6.44)$$

and with (6.36) for the temperature dependence of  $Q_{p0}$ , which results in

$$\zeta_{CT} = \zeta_{CT0} - \zeta_{Qp0}. \quad (6.45)$$

### 6.3.2.2 Back injection related base current components

The forward internal base current that is mostly determined by backinjection into the emitter was given by (3.101). The temperature dependent saturation current  $I_{BEIS}$  can be obtained from (3.99). There are two temperature dependent terms that will be considered separately below.

Inserting (6.5) and (6.11) into the first (diffusive) term of  $I_{BEIS}$  yields an expression similar to that of the transfer current,

$$\bar{\mu}_{pE} V_T n_{iE}^2(w_E) = \bar{\mu}_{pE}(T_0) V_{T0} n_{ie}^2(T_0) \left(\frac{T}{T_0}\right)^{\zeta_{BET}} \exp\left[\frac{V_{gEff}(T_0)}{V_{T0}} - \frac{V_{gEff}(T)}{V_T}\right] \quad (6.46)$$

with  $V_{gEff}$  as the *average effective* bandgap voltage in the relevant emitter region, which includes the (average) bandgap-narrowing voltage  $\Delta V_{gE}$  and its temperature dependence. Furthermore,

$$\zeta_{BET} = m_g + 1 - \zeta_{\mu pE} \quad (6.47)$$

with  $m_g = 3$  at this point. The value of the coefficient  $\zeta_{\mu pE}$  depends on the doping of the respective region. For conventional bipolar transistors and properly designed LEC HBTs, the relevant emitter doping is very high so that  $\zeta_{\mu pE}$  is fairly small (around 0.1).

Inserting (6.11) and (6.14) into the interface velocity recombination related term of (3.99) gives

$$1 + \frac{\bar{\mu}_{pE} V_T}{w_E v_{Epm}} = 1 + \frac{\mu_{pE}(T_0) V_{T0}}{w_E v_{pm}(T_0)} \left(\frac{T}{T_0}\right)^{1 - \zeta_{\mu pE} - 0.5}. \quad (6.48)$$

The temperature dependence of this expression is obviously fairly small. In the unlikely case of a very small value for  $v_{pm}$  the exponent ( $0.5 - \zeta_{\mu pE}$ ) can directly be subtracted from  $\zeta_{BET}$ . However, this case is not desirable in advanced processes, in which negligible interface oxide thickness is important to minimize the emitter resistance for very small emitter areas. On

the other hand, for mono-emitters or negligible influence of interfaces,  $v_{pm}$  represents the contact recombination velocity. In this case, (6.48) equals essentially 1.

Finally, inserting above results as well as the bandgap function (6.4) into (3.99) leads to the temperature dependent saturation current

$$I_{BEiS}(T) = I_{BEiS}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{BET}} \exp \left[ \frac{V_{gEff}(0)}{V_T} \left( \frac{T}{T_0} - 1 \right) \right], \quad (6.49)$$

where  $m_g$  in (6.47) is now given by (6.9). In practice,  $\zeta_{BET}$  may also contain a small contribution from the interface or contact recombination velocity related term (6.48). Note, that the factor  $m_{BEi}$  does not occur in (6.49) since it was introduced for taking into account *bias* dependence related simplifications. The same functional dependence can also be used for the injection related components of other transistor regions after replacing  $V_{gEff}$  and  $\zeta_{BET}$  by the parameters of the respective region. However, if the base current is controlled by tunneling through, e.g., an interfacial oxide, the temperature dependence can change significantly.

### 6.3.2.3 Recombination currents

The recombination current  $I_{REi}$  in the BE SCR is given by (3.112) and (3.113). The temperature dependence is determined mostly by  $\bar{n}_i$ , while the exact temperature dependence of  $\tau_{eq}$  is not well known but definitely very small compared to that of  $\bar{n}_i$ . The same is true also for T dependence of the SCR width  $w_{BE0}$ . Therefore, after inserting (6.5) along with the bandgap function (6.4) into (3.112) the recombination saturation current can be written as

$$I_{REiS}(T) = I_{REiS}(T_0) \left( \frac{T}{T_0} \right)^{(m_g/2)} \exp \left[ \frac{V_{gBEff}(0)}{2V_T} \left( \frac{T}{T_0} - 1 \right) \right], \quad (6.50)$$

with  $m_g$  from (6.9) and  $V_{gBEff}(0)$  as the average effective bandgap voltage in the BE SCR, which is also being used for the temperature dependence of the BE depletion capacitance. Note, that the factor  $m_{REi}$  does not occur in (6.50) since it was introduced for taking into account *bias* dependence

related simplifications. Any SCR recombination component related to other transistor regions can be described the same way.

### 6.3.2.4 Base-collector avalanche current

The classical local approximation (3.133) for the avalanche current contains as temperature dependent variables the transfer current  $I_T$ , the BC built-in voltage  $V_{DCi}$  (also in  $C_{jCi}$ ) and the ionization rate coefficients  $a_n$  and  $b_n$ . Inserting the temperature dependence of the latter from (6.13a,b) into (3.133) permits to calculate  $I_{AVL}$ . Due to the various temperature dependent terms in (3.133) and their different functional form it is generally very difficult to derive an expression of the form  $I_{AVL}(T) = I_{AVL}(T_0) g_{AVL}(T/T_0)$ .

In the non-local model, the avalanche current is calculated directly from (3.122) and the non-local multiplication factor (4.101). The temperature dependence is mainly given by the transfer current with its exponential dependence. A possible, but smaller impact can occur also from the coefficient  $b_{avm}$ , which is assumed to be directly proportional to  $b_n$  so that its  $T$  dependence reads

$$b_{avm}(T) = b_{avm}(T_0)[1 - \zeta_{bn}\Delta T]. \quad (6.50a)$$

Converting (6.13a) into a  $(T/T_0)^{-Z}$  form leads to the more convenient expression for the non-local avalanche current

$$I_{AVL}(T) \cong I_{AVL}(T_0) \frac{I_T(T)}{I_T(T_0)} \left( \frac{T}{T_0} \right)^{\zeta_{an}T_0} \exp(-b_{avm}\zeta_{bn}\Delta T). \quad (6.51)$$

### 6.3.2.5 Temperature dependence of current gain

The relative temperature coefficient (TC) of the current gain is often similar across process versions and even generations, and is also easy to determine from measurements. Hence, circuit designers are generally more interested in the specification of the temperature dependence of the *current gain* rather than of the base current. However, the "current gain" parameter used in certain compact models is not equivalent to the measured current gain at the terminals, but is defined for the internal transistor

only. To be precise, it is defined as the *extrapolated ideal forward internal* current gain towards  $V_{B'E'} = 0$ , which is designated here as  $B_{fi0}$ . The latter is given by the ratio of saturation current  $I_S(T)$  of the transfer current to the saturation current  $I_{BEiS}(T)$  of the BE backinjection current and reads after inserting (6.43) and (6.49)

$$B_{fi0}(T) = \frac{I_S(T_0) \left(\frac{T}{T_0}\right)^{\zeta_{CT}} \exp \left[ \frac{V_{gBEff}(0)}{V_T} \left(\frac{T}{T_0} - 1\right) \right]}{I_{BEiS}(T_0) \left(\frac{T}{T_0}\right)^{\zeta_{BET}} \exp \left[ \frac{V_{gEeff}(0)}{V_T} \left(\frac{T}{T_0} - 1\right) \right]}.$$

Sorting terms gives

$$B_{fi0}(T) = B_{fi00} \left(\frac{T}{T_0}\right)^{\zeta_{Bfi}} \exp \left[ \frac{\Delta V_{gBEeff}(0)}{V_T} \left(\frac{T}{T_0} - 1\right) \right] \quad (6.52)$$

with the low-injection current gain at reference temperature,

$$B_{fi00} = B_{fi0}(T_0) = I_S(T_0)/I_{BEiS}(T_0), \quad (6.53a)$$

the bandgap difference between neutral base and emitter,

$$\Delta V_{gBEeff}(0) = V_{gBEeff}(0) - V_{gEeff}(0), \quad (6.53b)$$

and a coefficient that is given by the difference

$$\zeta_{Bfi} = \zeta_{CT} - \zeta_{BET} \cong m_{\mu pE} - m_{\mu nB}. \quad (6.53c)$$

This general result for the ideal internal current gain together with the parameters in (6.53a,b,c) could be used directly to define the temperature dependence of the base current. Such an approach may be a compromise between modeling and circuit design simplicity. However, the temperature dependence of the current gain at different injection levels can be determined by different current components in both base and transfer current. Therefore, a generally applicable description that is as flexible as possible for industrial production use (i.e. simulator based design) and also from a model accuracy point of view is to describe each *current component separately* without trying to define their bias and temperature behavior from assumptions about the current gain. This will also provide a clear definition of how to extract the corresponding model parameters.

### 6.3.2.6 Tunnelling current

The temperature dependence of the parameters describing BE tunnelling is mainly determined by the temperature dependence of the bandgap both directly in the exponent of (4.102) and in the junction field (4.103). A useful expression for modeling can only be given though, once the exact model formulation has been defined. Since the tunneling current can be associated with the BE junction region, the bandgap voltage is approximated by the average value of the base and emitter bandgap voltage,

$$V_{gBEff}(T) = \frac{V_{gBEff}(T) + V_{gEEff}(T)}{2}, \quad (6.54)$$

which is already available from the calculation of the BE depletion capacitance and the BE recombination current.

### 6.3.3 Series resistances

The temperature dependence of the series resistances is determined mainly by the majority carrier mobility in the respective neutral region (via the corresponding sheet resistance) and, to a lesser extent, also by the built-in voltages (via the corresponding SCR widths). In contrast, the temperature dependent influence of contact resistances is negligible. Inserting the mobility equation (6.11) then leads to

$$R(T) = R(T_0) \left( \frac{T}{T_0} \right)^{\zeta_R} \quad (6.55)$$

with the parameter  $\zeta_R$  being a function of the (average) doping concentration of the respective region.

Conductivity modulation and emitter current crowding in  $R_{Bi}$  are automatically described as a function of  $T$  by the corresponding charges and currents. The temperature dependence of the substrate coupling resistance  $R_{SM}$  can be modeled with the same formulation as (6.55).

### 6.3.4 Noise

The temperature dependence of thermal and shot noise is given by the temperature behavior of the respective resistor and current values. In contrast, the flicker noise spectral density is described by an empirical expression (7.37) in which the parameters  $K_F$  and  $a_F$  may be a function of temperature. Experimental data for SiGe HBTs in [19, 20] show  $a_F \approx 2$  and a decrease of the spectral density with  $T$ . The results indicate a linear decrease of  $\log(k_F)$  with temperature, leading to

$$K_F(T) = K_F(T_0) \exp(-\zeta_{KF} \Delta T)$$

with the temperature coefficient  $\zeta_{KF} \approx 0.012/\text{K}$  for npn HBTs and  $\zeta_{KF} \approx 0.024/\text{K}$  for pnp HBTs [20].

### 6.4 Self-heating effects

As mentioned earlier, the power dissipated within the device structure leads to a temperature increase. In bipolar transistors, the highest potential drop exists in the BC SCR. Together with the transfer current or, in general, the in-phase component of the collector current, this leads to heat generation. The corresponding heat flow is determined by the spatial extension, material composition, and boundaries of the overall device structure. In general, this is a 3D modeling problem in space and time. The corresponding governing equations describe the heat flow,

$$J_h = -\kappa(T) \text{grad} T, \quad (6.56)$$

and the continuity (or balance) of heat flow and exchange,

$$\text{div} J_h = p - c_h \frac{\partial T}{\partial t}. \quad (6.57)$$

Here,  $T$  is the temperature in K;  $p$  is the power density (in  $\text{W}/\text{cm}^3$ ) generated by heat sources;  $c_h$  is the specific thermal capacitance density (in  $\text{J}/(\text{Kcm}^3)$ ) that is defined by the product of specific mass density and specific heat capacity of the material;  $\kappa$  is the temperature dependent thermal conductivity (in  $\text{W}/(\text{Kcm})$ ),



$$\kappa(T) = \kappa_0 \left( \frac{T}{T_0} \right)^{-\alpha}. \quad (6.58)$$

$\kappa_0 = \kappa(T_0)$  and  $\alpha$  are material dependent parameters the values of which are listed in Table 6.2. According to [21, 22], the above equation is accurate for silicon within 3% over the temperature interval [250, 1000]K. The solution of the resulting 3D nonlinear partial differential equation system for a given set of boundary conditions yields the temperature distribution  $T(x, y, z, t)$ .

material	$\kappa_0$ (WK <sup>-1</sup> cm <sup>-1</sup> )	$\alpha$	$c_h$ (JK <sup>-1</sup> cm <sup>-3</sup> )
Si	1.702	1.4	1.637
Ge	0.6	1.25	1.92
InP	0.658	1.48	1.97
GaAs	0.4	1.27	1.71
SiO <sub>2</sub>	0.014	0.33	1.7... 2.07
Si <sub>3</sub> N <sub>4</sub>	0.185	0.33	2.44

Table 6.2: Thermal parameters at  $T_0 = 300$ K [4,22] for selected HBT related materials.

Today, various tools exist for solving the above PDE in a rigorous numerical way (e.g. [23, 24]). These solutions provide valuable insight into the details of heat flow and temperature distribution in realistic device structures and can serve as reference for simplified models. However, they are quite time consuming to set up and computationally too expensive for direct use in or with compact models. Also, a detailed distributed temperature is generally not a suitable input variable for circuit simulation. Therefore, electrical equivalents of the PDE and its solution have been developed by the semiconductor community already early on (e.g. [25]). This is achieved by the following transformations: power  $P \rightarrow$  current  $I$ , temperature  $T \rightarrow$  potential  $V$ , thermal conductivity  $\rightarrow$  electrical conductivity, and thermal capacitance  $\rightarrow$  electrical capacitance.

Below, those solutions will be briefly discussed that are suitable for circuit simulation applications. Since the large volume of publications on various methods for solving the heat PDE would fill a book on its own (e.g. [26]), and since for compact HBT modeling only an as simple as possible compact equivalent circuit is desired, the focus will be on a concise summary of methods that can be employed to efficiently calculate the elements of a compact thermal equivalent circuit as a function of the relevant features of the device structure.

#### 6.4.1 Intra-device electro-thermal coupling

The most simple case to start with is the steady-state temperature distribution. With  $\partial T/\partial t = 0$  still a non-linear differential equation exists. However, by performing the Kirchhoff transformation [27],

$$\Theta = T_s + \frac{1}{\kappa(T_s)} \int_{T_s}^{T(x,y,z)} \kappa(T') dT', \quad (6.59)$$

the variable  $T$  is replaced by  $\Theta$ , so that the steady-state version of (6.56) and (6.57) can be linearized yielding simply a Poisson-type equation

$$\text{div}(\kappa(T) \text{grad} T) = \kappa_0 \text{div} \text{grad} \Theta = -p. \quad (6.60)$$

$T_s$  is here the spatially independent substrate (or chip) temperature at the backside of the wafer. Once a solution of (6.60) for  $\Theta$  is known the actual temperature can be calculated from solving (6.59) for  $T(x,y,z)$ . Note that (6.60) is only valid for homogeneous material. In a multi-layer or multi-material structure the solutions for each layer would have to be connected to each other using appropriate interface conditions. Analytical solutions for (6.60) exist for simple homogeneous rectangular structures (e.g. [26, 28]). According to [29, 30], such solutions can be extended to multi-layer structures by, e.g., the method of images.

For compact modeling the temperature distribution needs to be reduced to preferably a scalar variable representing, e.g., the average device temperature (cf. Fig. 6.4). A convenient approach is to define a thermal resistance  $R_{th}$  such that, according to the electro-thermal analogy mentioned earlier, the product  $R_{th}$  times total power  $P$  dissipated in the device equals

the temperature increase at a selected location  $(x_0, y_0, z_0)$  in the device with respect to a known reference temperature  $T_s$ :

$$R_{th}(x, y, z, T_s, P) = \frac{T(x, y, z) - T_s}{P} = \int_{l(x_0, y_0, z_0)}^{l(x, y, z)} \frac{\kappa}{A(l')} dl' \quad (6.61)$$

The latter expression corresponds directly to a “resistance” definition from the heat flux (current) path between two points.

For the device temperature, usually the BE junction is selected since it controls the most important device characteristics. In other words, the thermal voltage in the equivalent circuit element equations corresponds to this junction temperature  $T_j$ . With these definitions electro-thermal simulation consists basically of a two-step iterative procedure. In the first step the electrical characteristics are evaluated at  $T_j$ . Then the results are used to calculate the dissipated power  $P$  and the updated junction (device) temperature

$$T_j = T_s + R_{th}P. \quad (6.62)$$

A rough estimate is  $P \approx I_C V_{CE}$ , while in compact models all dissipative terms can be easily included in the calculation.

As suggested in [21], the two-step solution procedure of the heat equation can be employed for an efficient formulation of the thermal resistance that includes the temperature dependence of the thermal conductivity. First, the thermal resistance at “low power” levels,  $R_{th0}$ , is calculated for the thermal conductivity given at the substrate temperature  $T_s$ ,

$$R_{th0}(T_s) = \frac{\Theta - T_s}{P} = R_{th0}(T_0) \left[ 1 + \alpha \left( \frac{T_s}{T_0} - 1 \right) \right], \quad (6.63)$$

with  $R_{th0}(T_0)$  as the thermal resistance at the (parameter) reference temperature. Eq. (6.63) involves a linearization of  $1/\kappa(T)$  at  $T_0 = 300\text{K}$ , the maximum error of which is about 5% at 150 K increase beyond  $T_0$ .

In the second step, the junction temperature is calculated from the evaluation of (6.59) using (6.58) and then inserting  $\Theta$  from the middle expression of (6.63), yielding

$$T_j = T_s \left[ 1 + \frac{(1 - \alpha) R_{th0}(T_s)}{T_s} P \right]^{\frac{1}{1 - \alpha}}. \quad (6.64)$$

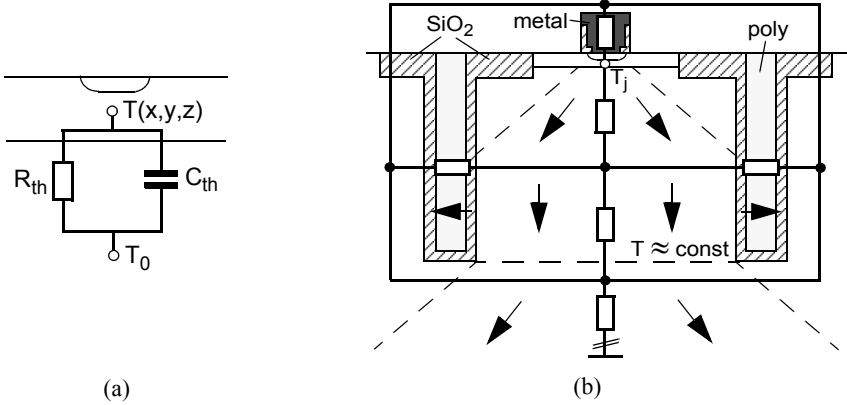


Fig. 6.4: Illustration of thermal resistance related self-heating. (a) General electrical equivalent circuit with thermal resistance  $R_{th}$  and thermal capacitance  $C_{th}$ . (b) Representation of thermal resistance network in a DTI structure; arrows indicate heat flux in the substrate.

The final value of the thermal resistance is then

$$R_{th}(T_j) = \frac{T_j - T_s}{P} \approx R_{th0}(T_s) \left[ 1 + \frac{\alpha R_{th0}(T_s)}{2 T_s} P \right], \quad (6.65)$$

The last expression results from a linearization of (6.64) and was shown in [21] to have deviations lower than 5% for a temperature increase of 100K in InP, which has one of the highest values of  $\alpha$ .

According to (6.61)  $R_{th}$  can be calculated as function of geometry from the temperature distribution of a given device structure. Besides the already mentioned time consuming rigorous numerical solution there are various other methods for calculating  $T(x, y, z)$ . Generally, their speed comes at the expense of reduced accuracy due to necessary simplifications. One can distinguish between *series-based* and *compact* solutions.

Early attempts at solving the heat equation analytically for transistor modeling assumed a semi-infinite volume for 3D heat spread with a fixed reference temperature at wafer backside. A Green's function based ap-

proach combined with the method of images (e.g. [31]) or separation of variables (e.g. [32, 33]) provided a solution for  $T(x,y,z)$  and allowed to calculate  $R_{th}$  explicitly as a function of emitter dimensions. While in [33] a rectangular (2D) heat source at the semiconductor *surface* was assumed, the same concept was in [31, 32, 28] applied to the more realistic case of a rectangular *volumetric* heat source. Nevertheless, such results are valid only for larger structures and junction isolation.

Advanced SiGe BiCMOS technologies utilize deep trench isolation (DTI) as sketched in Fig. 6.4. With trench depths around  $4\ldots 7\text{ }\mu\text{m}$ , which is much larger than the BC junction depth, heat spread in lateral direction is very limited compared to vertical direction. This anisotropic heat flow requires more sophisticated calculation methods for  $R_{th}$ . Initial investigations of DTI SiGe HBT structures using a volumetric heat source revealed an increase of the thermal resistance by over 60% compared to junction isolation, thus making the accurate representation of self-heating even more important [34, 32, 35]. By performing numerical solutions as reference, it was also shown in [34, 35] that the temperature across the bottom of the DTI box was almost constant. This allows a breakup of the  $R_{th}$  calculation into the trench box and the classical semi-infinite region.

Further improvement of model accuracy was achieved by taking into account both lateral heat flow across the trench walls and vertical heat flow upwards through the emitter metallization [32, 36]. The DTI heat flow was included via a boundary condition of the form  $dT/dx = HT$ , with  $H$  as DTI heat transfer coefficient, and an Green's function based solution with improved computational efficiency compared to the method of images. The thermal resistance of the emitter metallization was calculated in [36] from simple lumped considerations. This improved model exhibited excellent agreement for  $R_{th}$ , especially for short structures, with both experimental data and purely numerical solutions.

Based on insight, such as a constant temperature at the bottom of the DTI box, from both semi-numerical and rigorous numerical solutions, compact models for simplified but practical structures emerged (e.g. [37, 39]), which include a non-symmetric location of the heat source w.r.t. the surrounding trench walls and heat flow through the trench walls [39].

The considerations above are based on the steady-state condition. During actual transistor operation the transient response of the heat flow may

come into play. The typical local thermal time constant, i.e. the time for reaching steady-state in response to a step of power dissipation change in single device, is in the order of  $0.1 \dots 1 \mu\text{s}$  (e.g. [31, 40, 41]). Although this is a much slower process than changes in electrical variables, dynamic electro-thermal effects will matter in those cases in which transistors spend a long time at a certain bias point before they are switched into a different bias point. Once the idle time comes into the order of 100 ns the temperature change may change significantly and have a noticeable impact on the transient behavior (depending on the dissipated power). This can cause jitter if the idle time varies randomly as, e.g., in laser drivers.

Solving the linear heat equation in time or frequency domain yields for the temperature an infinite series of  $\exp(-t/\tau_n)$  terms with  $n = 1 \dots \infty$  and  $\tau_n < \tau_{n-1}$ . According to (6.61) this solution can be transformed into a thermal *impedance* that can be represented by a series connection of an infinite number of RC networks with different time constants (eg. [26, 42]). However, for compact modeling only a limited number of terms, i.e. RC sub-circuit *nodes*, can be included for computational reasons. In larger circuits very often only a single-pole thermal network is feasible (cf. Fig. 6.4). As was shown in [41, 43] though it takes generally three terms, i.e. time constants, to approximate the actual transient thermal response with sufficient accuracy. Small-signal considerations of the electro-thermal effect in device characteristics can be found in [44, 45].

The electro-thermal feedback can lead to instability of transistor behavior depending on the biasing scheme. This shall be discussed briefly since it has significant impact on transistor measurements and parameter extraction. It is assumed that  $V_{CE}$  is kept at a constant value and that self-heating already occurs at the initial bias point ( $I_{C0}$ ,  $V_{CE}$ ), while the avalanche current can still be neglected. Furthermore, the DC collector current is approximated according to (6.42) by its most temperature sensitive (exponential) term as:

$$I_C = I_0 \exp\left(\frac{V_{BE} - V_{gB0}}{V_T}\right) \quad (6.66)$$

with  $V_{gB0} = V_{gB\text{eff}}(0)$ . Due to  $V_{BE} - V_{gB0} < 0$  the temperature independent current  $I_0 = I_S \exp(V_{gB0}/V_{T0})$  is larger than  $I_C$ .

For constant  $V_{BE}$  the temperature increase due to self-heating leads to a larger thermal voltage and, hence, a smaller absolute value of the argument of the exp-function, which in turn increases. At constant  $V_{CE}$  the dissipated power increases and so does the temperature. This is a positive feedback effect (also known as thermal runaway) that in practice is limited by series resistances of mainly base and emitter the impact of which also occurs in the exponent. Often additional external resistances, so-called ballast resistances, are added to make the device thermally stable at the cost of reduced speed due to increased electrical RC time constants.

For constant base current the collector current increases (decreases) if the current gain temperature coefficient (TC) is positive (negative). The sign and value of the latter is determined, according to (6.52), by the band-gap difference  $\Delta V_{gBEff}(0)$  between emitter and base region (cf. (6.53b)). For BJTs  $\Delta V_{gBEff}(0) > 0$  causing thermal runaway that can be limited by an emitter series and ballast resistance. In contrast,  $\Delta V_{gBEff}(0) \leq 0$  in SiGe and III/V HBTs leads to a negative current gain TC and, thus, initially to a thermally more stable operation at constant  $I_B$ . However, for high power dissipation the so-called collector current collapse occurs.

The exact point of thermal instability obviously depends on the transistor operation mode and the influence of series resistances as well as, according to the (G)ICCR, the temperature dependent behavior of the charges. There exists plenty of literature dealing with this topic and attempting, e.g., analytical solutions with the goal of deriving criteria for choosing ballast resistance values, for properly spacing emitter fingers, or for device layout that improves thermal stability [28, 33, 42, 46, 47, 48].

#### 6.4.2 Inter-device electro-thermal coupling

The spread of heat flux causes a temperature rise also in the region around the one generating the heat. As shown in Fig. 6.5a this can impact both the emitter fingers within the same (multi-finger) transistor and the adjacent devices. The result is a temperature dependent shift of the associated characteristics of these device structures. Within a multi-emitter transistor each emitter contributes to power and heat generation, thus leading to an inhomogeneous temperature distribution as shown in Fig. 6.5b for a five finger structure. The inhomogeneity eventually causes the hottest fin-

ger to hog the entire current which can destroy the device. As a consequence, electro-thermal coupling needs to be taken into account during circuit design.

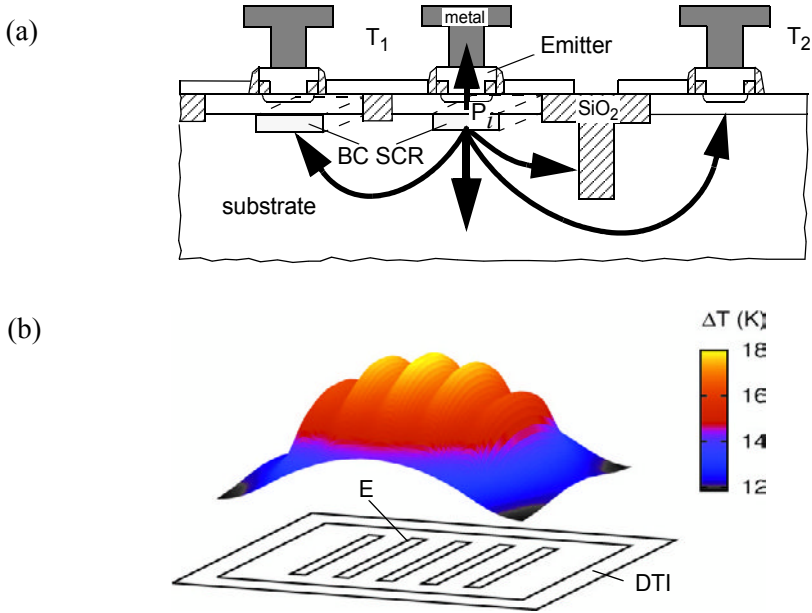


Fig. 6.5: (a) Illustration of heat flux (indicated by arrows) and resulting thermal coupling between different regions within a transistor and also due to heat spread to adjacent devices. (b) Calculated temperature distribution in a 5-finger transistor. The projection shows the finger and trench layout [43].

The heat flow and temperature distribution can be calculated from either full numerical or series-based solutions of (6.56) and (6.57). This information then needs to be converted to a form that is suitable for circuit simulation and design using compact models. The most straightforward approach for simulating the coupling between different devices is to use the already defined junction (device) temperature from the intra-device self-heating effect in the previous section and then to add the temperature increase that results from the impact of the adjacent device (or devices). Since the temperature increase  $\Delta T_k$  of device  $k$  consists of a superposition of the temperature increase caused by its own self-heating and that of all adjacent devices  $j = 1 \dots K, j \neq k$ , the impact of each adjacent device can be



calculated separately from solving (6.56) and (6.57) successively with only the heat source  $P_j$  turned on. This gives the temperature increase  $\Delta T_{kj}$  at device  $k$  (and also the increase at all other devices of interest). These devices are coupled by the thermal resistance  $R_{th,kj} = \Delta T_{kj}/P_j$ . The corresponding thermal coupling coefficient between device  $k$  and  $j$  is

$$c_{kj} = R_{th,kj}/R_{th,j} \quad (6.67)$$

with  $R_{th,j}$  as the thermal self-resistance of device  $j$ .

(6.67) along with the intra-device electro-thermal circuit modeling approach allows now to construct an electrical equivalent circuit for thermal coupling between devices that minimizes the number of nodes and, hence, the computational effort [49, 43]. Figure 6.6a shows an example of such an equivalent circuit for two transistors. The self-heating is included the same way as in the previous section using the “self”-thermal resistance, while the coupling is realized by a voltage controlled voltage source (VCVS) in series to the “self”-thermal resistance. The respective temperature increase for each device reads

$$\begin{aligned} \Delta T_1 &= R_{th,1}P_1 + c_{12}\Delta T_2 \\ \Delta T_2 &= c_{21}\Delta T_1 + R_{th,2}P_2 \end{aligned} \quad (6.68)$$

This concept can be easily extended to multiple devices and heat sources as shown in Fig. 6.6b. The VCVS then includes the contribution of all other relevant heat sources. The practical realization of this concept requires access to the thermal node of each device and turning off the impact of its model parameter  $R_{th}$ . The same idea also applies to coupling between emitters of the same device. In this case, each emitter needs to be represented by a separate transistor (model) and corresponding heat source. In practice, series-based solutions of the heat equation are the most efficient way to calculate the coupling coefficients as compared to full numerical solutions.

The approach described above can be extended further to the case of time dependent coupling by properly taking into account the delayed impact of an adjacent heat source ( $j$ ) on the heat source of interest ( $k$ ). For a single-pole (i.e. single time constant) approximation, the time dependent temperature increase at device  $k$  reads

$$\Delta T_{kj}(t) = c_{kj} \Delta T_{jj}(t) \left[ 1 - \exp\left(\frac{-t}{\tau_{th, kj}}\right) \right]. \quad (6.69)$$

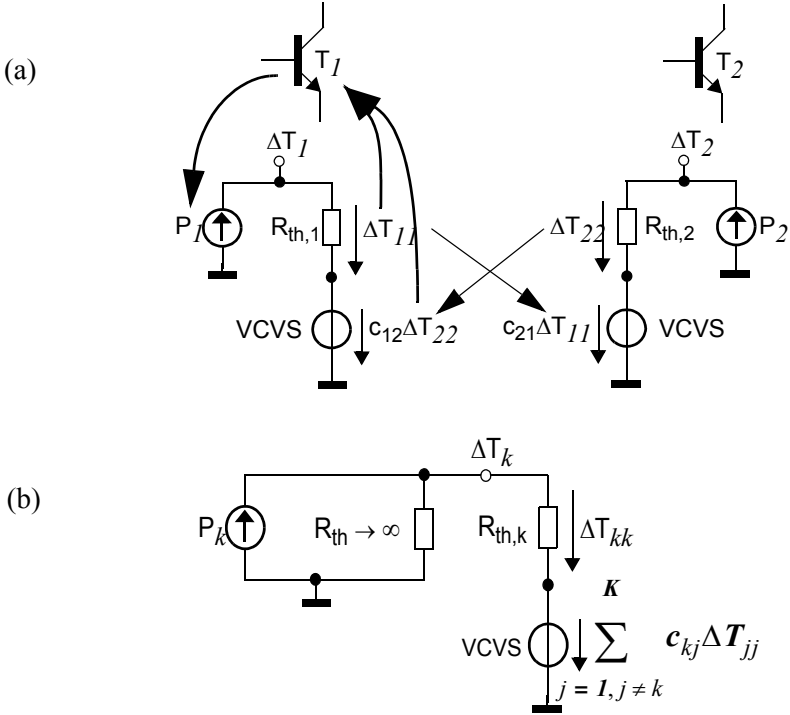


Fig. 6.6: Electrical representation of thermal coupling: (a) Between two heat sources, such as transistors or fingers. (b) Generic representation in a selected device (k) for coupling from multiple ( $K-1$ ) heat sources.

with the thermal coupling time constant  $\tau_{th, kj} = R_{th, kj} C_{th, kj}$ . For a network representation the delay can be realized simply by a low-pass circuit with the thermal resistance  $R_{th, kj}$  and thermal capacitance  $C_{th, kj}$  related to the coupling of the devices. Figure 6.7 exhibits the complete electrical network representation for transient electro-thermal coupling. The additional time-response related network consists of the attenuated temperature  $c_{kj} \Delta T_{jj}(t)$  that feeds as voltage controlled voltage source (VCVS) in the delay network. The coupling time constant can be calculated from (or fitted

to) the time dependent solution of the heat equation with a series-based method being the most efficient approach.

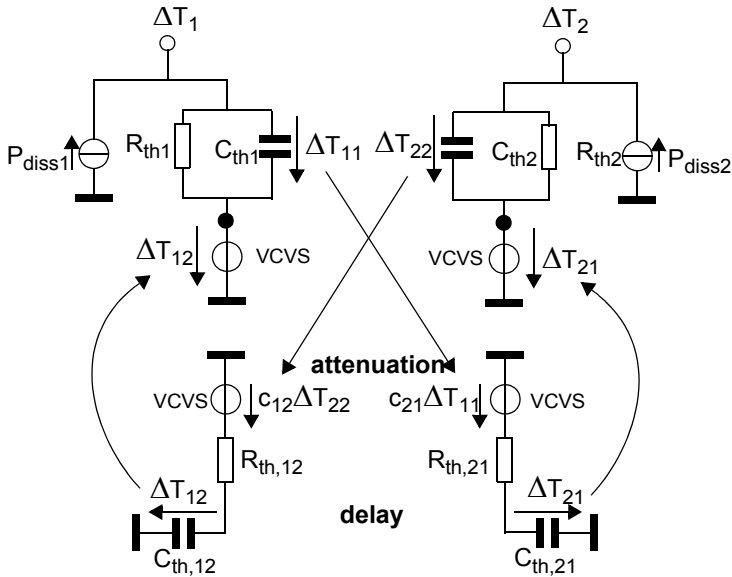


Fig. 6.7: Electrical representation of transient thermal coupling between two heat sources, i.e. transistors or fingers [43].

## 6.5 References

- [1] K. Brennan, "The physics of semiconductors", Cambridge University Press, 1999.
- [2] A. Schenk, "Advanced physical models for silicon device simulation", Springer Wien New York, 1998.
- [3] C. D. Thurmond, "The standard thermodynamic function of the formation of electrons and holes in Ge, Si, GaAs and GaP", J. Electrochem. Soc., Vol. 122, p. 1133, 1975.
- [4] S. Sze, "Physics of semiconductor devices", Wiley & Sons, New York, 1981.
- [5] S. Lin and C. Salama, "A  $V_{BE}(T)$  model with application to band-gap reference design", IEEE Journal of Solid-State Circuits, Vol. 20, pp. 1283-1285, 1985.
- [6] D. Celi, private communications, 2004.

- [7] E.O. Kane, "Theory of tunneling", J. App. Phys., Vol. 32, No. 1, pp. 83-92, 1961.
- [8] J.W. Slotboom and H.C. deGraaff, "Measurements of bandgap narrowing in Si bipolar transistors", Solid-State Electronics, Vol. 19, pp. 857-862, 1976.
- [9] M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors with special regard to high current densities and two-dimensional effects", Ph. D. thesis (in German), Ruhr-University Bochum, Bochum, Germany, 1988.
- [10] C. Jacoboni et al., "A review of some charge transport properties of silicon", Solid-State Electronics, Vol. 20, pp. 77-89, 1977.
- [11] R. Ranfft, "Investigation and optimization of subnanosecond bipolar circuits for highly integrated building blocks", Dissertation (in German), Ruhr-University Bochum, 1981
- [12] W. Maes, K. DeMeyer, and R. Van Overstraaten, "Impact ionization in silicon: a review and update", Solid-State Electron., Vol. 33, pp. 705-718, 1990.
- [13] V. Kunz, C. deGroot, S. Hall, and P. Ashburn., "Polycrystalline Silicon-Germanium emitters for gain control, with application to SiGe HBTs", IEEE Trans. Electron Dev., Vol. 50, pp. 1480-1486, 2003.
- [14] Q. Liang, R. Krithivasan, A Ahmed, Y. Lu, Y. Li. J. Cressler, G. Niu, J.-s. Rieh, G. Freeman, D. Ahlgren, A Joseph, "Analysis and understanding of unique cryogenic phenomena in state-of-the-art SiGe HBTs", Solid-State Electronics, Vol. 50, pp. 964-972, 2006.
- [15] J. Cressler (ed.), "Silicon heterostructure Handbook", CRC Press NY, 2005.
- [16] D. Richey, A. Joseph, J. Cressler, R. Jaeger, "Evidence for non-equilibrium base transport in Si and SiGe bipolar transistors at cryogenic temperatures", Solid-State Electronics, Vol. 39, pp. 785-789, 1996.
- [17] C. McAndrew, private communication, 2002.
- [18] M. Schröter and T.-Y. Lee., "A physics-based minority charge and transit time model for bipolar transistors", IEEE Trans. Electron Dev., vol. 46, pp. 288-300, 1999.
- [19] T. Chen, W.-M. Kuo, E. Zhao, Q. Liang, Z. Jin, J. Cressler, A. Joseph, "On the high-temperature (to 300 °C) characteristics of SiGe HBTs", IEEE Trans. Electron Dev. Vol. 51, pp. 1825-1832, 2004.

- [20] E. Zhao, R. Krithivasan, A. Sutton, Z. Jin, J. Cressler, B. El-Kareh, S. Balster, H. Yasuda, "An Investigation of low-frequency noise in complementary SiGe HBTs", *IEEE Trans. Electron Devices*, Vol. 53, No.2, pp. 329-338, 2006.
- [21] D.J. Walkey, T. Smy, T. Macelwee, M. Maliepaard, "Compact representation of temperature and power dependence of thermal resistance in Si, InP, GaAs substrate devices using linear models", *Solid-State Electronics*, Vol. 46, pp. 819-826, 2002.
- [22] V. Palankowski, R. Quay, "Analysis and Simulation of heterostructure devices", Springer, Wien New York, 2004.
- [23] T.J. Smy, D.J. Walkey, S. K. Dew, "A 3D thermal simulation tool for integrated devices-Atar", *IEEE Trans. on CAD of Integrated Circuits and Systems*, Vol. 20, pp. 105-115, 2001.
- [24] *ATLAS User's Manual*, Silvaco Int., Santa Clara, 2002.
- [25] P. Strickland, "The Thermal Equivalent Circuit of a Transistor", *IBM J. of Res. Dev.*, Vol. 3, No. 1, p. 35, 1959.
- [26] M. Özisik, "Heat conduction", J. Wiley & Sons, NY, ISBN 0-471-53256-8, 1993.
- [27] W. Joyce, "Thermal resistance of heat sinks with temperature dependent conductivity", *Solid-State Electron.*, Vol. 18, pp. 321-322, 1975.
- [28] N. Rinaldi, "Thermal analysis of solid-state devices and circuits: an analytical approach", *Solid-State Electronics*, Vol. 44, pp. 1789-1798, 2000.
- [29] P. Leturq et al., "A new approach to thermal analysis of power devices", *IEEE Trans. Electron Dev.*, Vol. 34, pp. 1147-1156, 1987.
- [30] A. Palisoc, C. Lee, "Exact thermal representation of multilayer rectangular structures by infinite plate structures using the method of images", *J. App. Phys.*, Vol. 64, pp. 6851-6857.
- [31] R. Joy and E. Schlig, "Thermal properties of very fast transistors", *IEEE Trans. Electron Dev.*, Vol. 17, pp. 586-594, 1970.
- [32] J. Brodsky, "Physics-based thermal impedance models for the simulation of self-heating in semiconductor devices and circuits", Ph.D. thesis, University of Florida, Gainesville, 1997.
- [33] W. Liu, "Handbook of III-V heterojunction bipolar transistors", Wiley, New York, 1998.
- [34] D.J. Walkey, T.J. Smy, H. Tran, D. Marchesan, M. Schröter, "Prediction of thermal resistance in trench isolated bipolar device struc-

- tures", Proc. IEEE Bipolar and BiCMOS Circuits and Technology Meeting, Minneapolis 12.3, pp. 207-208, 1998.
- [35] D.J. Walkey, T.J. Smy, D. Marchesan, H. Tran, M. Schröter, "A scalable thermal model for trench-isolated bipolar devices", *Solid-State Electronics*, Vol. 44, No. 8, pp. 1373-1380, 2000.
  - [36] D.J. Walkey, T.J. Smy, C. Reimer, M. Schröter, H. Tran, D. Marchesan, M. Jackson, and T. Kleckner, "Modeling thermal resistance in trench-isolated bipolar technologies including trench heat flow", *Solid-State Electronics*, Vol. 46, No. 1, pp. 7-17, 2002.
  - [37] A. Pacelli, P. Palestri, M. Mastrapasqua, "Compact modeling of thermal resistance in bipolar transistors on bulk and SOI substrates", *IEEE Trans. Electron Dev.*, Vol. 49, pp. 1027-1033, 2002.
  - [38] N. Rinaldi, "On the modeling of the transient thermal behavior of semiconductor devices", *IEEE Trans. Electron Dev.*, Vol. 48, pp. 2796-2802, 2001.
  - [39] T. Vanhoucke, G.A.M. Hurkx, "A new analytical model for the thermal resistance of deep-trench bipolar transistors", *IEEE Trans. Electron Devices*, Vol. 53, pp. 1379 - 1388, 2006.
  - [40] R. Dennison and K. Walter, "Local thermal effects in high performance bipolar devices/circuits", Proc. IEEE Bipolar and BiCMOS Circuits and Technology Meeting, Minneapolis, pp. 164-167, 1989.
  - [41] D.J. Walkey, T.J. Smy, D. Marchesan, H. Tran, C. Reimer, T.C. Kleckner, M.K. Jackson, M. Schröter and J.R. Long, "Extraction and modeling of thermal behavior in trench isolated bipolar structures", *IEEE Bipolar and BiCMOS Circuits and Technology Meeting*, Minneapolis, pp. 97-100, 1999.
  - [42] R. Anholt, "Electrical and thermal characterization of MESFETs, HEMTs, and BJTs", Artech House, Boston, 1995.
  - [43] Y. Zimmermann, "Modeling of spatially distributed and sizing effects in high-performance bipolar transistors", MSEE/Diploma thesis, Chair for Electron Devices and Integrated Circuits, TU Dresden, 2004.
  - [44] R. Fox and S. Lee, "Scalable small-signal model for BJT self-heating", *IEEE Electron Dev. Lett.*, Vol. 12, pp. 649-651, 1991.
  - [45] N. Rinaldi, "Small-signal operation of semiconductor devices including self-heating with application to thermal characterization and instability analysis", *IEEE Trans. Electron Dev.*, Vol. 48, pp. 323-331, 2001.

- [46] C.-H. Liao, C-P. Lee, N. Wang, "Optimum design for a thermally stable multifinger power transistor", *IEEE Trans. Electron Devices*, Vol. 49, pp. 902-908, 2002.
- [47] C.-H. Liao, C-P. Lee, "Optimum design for a thermally stable multifinger power transistor with temperature dependent thermal conductivity", *IEEE Trans. Electron Devices*, Vol. 49, pp. 909-915, 2002.
- [48] D.J. Walkey, D. Celo, T.J. Smy, R.K. Surridge, "A thermal design methodology for multifinger bipolar transistor structures", *IEEE Transactions on Electron Devices*, Vol. 49, pp. 1375-1383, 2002.
- [49] D.J. Walkey, T. Smy, R. Dickson, J. Brodsky, D. Zweidinger, R. Fox, "Equivalent circuit modeling of static substrate thermal coupling using VCVS representation", *IEEE J. Solid-State Circuits*, Vol. 37, pp. 1198-1206, 2002.

This page is intentionally left blank



## **Chapter 7**

# **Compact Noise Modeling**

## 7.1 Basic noise mechanisms

The purpose of this chapter is to equip the reader with the minimum necessary foundation for noise modeling by providing a brief review on the basic noise mechanisms in semiconductors. This “noise primer” can by no means be a comprehensive discussion as the topic in itself fills entire books. For detailed derivations and discussions the reader is referred to [1-6] and the literature therein.

### 7.1.1 Fundamentals

Noise theory is based on the mathematical foundation of stochastic processes. Most books therefore dedicate at their beginning a fair amount of text to definitions, statistics, and abstract mathematical proofs (e.g. [1-5]) but unfortunately often do not provide explanations for the physical meaning of the mathematical expressions, especially for semiconductor devices. A simple example is the autocorrelation function which is usually just *defined* without any further comment. This does not necessarily contribute to making the realm of noise modeling attractive for (electrical) engineering students. While the theoretical foundations and mathematical proofs are certainly required for putting noise theory and research on solid grounds, they are supposed to “just” establish the foundation and methods from which engineering solutions to *practical* problems can be derived. Hence, this has to be the starting point for this chapter. Below, a concise summary of the most important concepts for modeling noise in semiconductors will be given with an attempt to provide some electrical engineering relation to the relevant abstract theorems. The corresponding derivations are not claimed to be mathematically exact but rather serve for illustrating the physical background or the application of a method.

The noise considered here is assumed to be a phenomenon in which the current  $i(t)$  through or the voltage  $v(t)$  across a semiconductor region fluctuates over time  $t$  around its DC value. Be  $u = \{i, v\}$  a stochastic variable then the DC value is defined as the mean (or average) of the stochastic signal,

$$\langle u(t) \rangle = \int_{-\infty}^{\infty} u(t) f_u du, \quad (7.1)$$

which is often also called first moment or expectation value. Here,  $f_u(u)$  is the probability of occurrence for the value  $u$ , so that  $f_u$  is called probability distribution function. In the following text,  $\langle . \rangle$  will be used synonymously for the integral expression.

Since noise is concerned with the fluctuation around  $\langle u \rangle$  it is convenient to introduce the fluctuating variable

$$\Delta u(t) = u(t) - \langle u \rangle. \quad (7.2)$$

The goal is to obtain a description of  $\Delta u(t)$  that contains the underlying physical mechanisms in semiconductors and is suitable for network analysis. Since  $\Delta u$  varies randomly around  $\langle u \rangle$  its mean value is zero and thus does not contain information relevant to noise. However, the variance (i.e. the second moment) defined as

$$\sigma_u^2 = \langle \Delta u(t)^2 \rangle = \int_{-\infty}^{\infty} \Delta u(t)^2 f_{\Delta u} du \geq 0 \quad (7.3)$$

appears to be a better choice. Note that the square of the amplitude is proportional to the power dissipated by the signal.

Another definition using the square of the signal is its *time average*

$$\overline{\Delta u(t)^2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \Delta u(t)^2 dt \quad (7.4)$$

taken over a sufficiently long period of time,  $T$ , which obviously must be finite for practical measurements. In practice, mostly *stationary* stochastic processes occur and are taken into account for circuit design. Of particular interest is the class of ergodic processes, which are defined by the *ensemble* mean values being equal to their *time* mean values. In other words, the average value of a desired property (e.g. velocity) for an ensemble (e.g. carriers) at a given time is equal to the average value of this property for a single ensemble member averaged over time. This gives for the variance

$$\sigma_u^2 = \langle \Delta u(t)^2 \rangle = \overline{\Delta u(t)^2}. \quad (7.5)$$

In electron devices and circuits noise is a small-signal phenomenon. Thus, it appears to be more convenient to treat its impact in frequency do-

main rather than in time domain. Fourier transform of the noise signal  $\Delta u(t)$  leads to the phasor

$$\Delta \underline{U}(\omega) = \hat{U} \exp(j\omega t) = \int_{-T}^T \Delta u(t) \exp(-j\omega t) dt, \quad (7.6)$$

assuming that the signal is zero outside of  $[-T, T]$  and that  $T$  is sufficiently large. Fundamentally,  $T$  has to be finite since otherwise the energy (given by the integral in (7.4)) can become infinite, although the power of the stochastic process itself is finite. Therefore, measuring noise corresponds to measuring power (or spectral power density to be more precise) as will become evident below.

The (noise related) root mean square (r.m.s) power in a resistor is proportional to

$$\Delta \underline{U} \Delta \underline{U}^* = \int_{-T}^T \int_{-T}^T \Delta u(t_1) \Delta u(t_2) \exp(-j\omega(t_2 - t_1)) dt_1 dt_2 \quad (7.7)$$

where  $\underline{U}^*$  is the conjugate complex of  $\underline{U}$ . Taking the mean of this expression and recognizing  $|\Delta \underline{U}|^2 = \Delta \underline{U} \Delta \underline{U}^*$  leads to

$$\langle |\Delta \underline{U}|^2 \rangle = \int_{-T}^T \int_{-T}^T \langle \Delta u(t_1) \Delta u(t_2) \rangle \exp(-j\omega(t_2 - t_1)) dt_1 dt_2, \quad (7.8)$$

where the mean integration (similar to (7.1)) on the right hand side (r.h.s.) was interchanged with the time integration and the  $\exp(\dots)$  function was left outside the mean expression since it does not depend on  $u$ . Introducing  $\tau = t_2 - t_1$  and  $t = t_1$  gives for (7.8)

$$\langle |\Delta \underline{U}|^2 \rangle = \int_{-T}^T \int_{-T}^T R_u(t, \tau) \exp(-j\omega\tau) dt d\tau, \quad (7.9)$$

with the so-called autocorrelation function [6]

$$R_u(t_1, \tau) = \langle \Delta u(t_1) \Delta u(t_1 + \tau) \rangle. \quad (7.10)$$

This function is a measure of the memory of a random process. It is typically only non-zero within a characteristic time constant (such as a scattering time) around the point in time at which such an event occurs (cf. Fig. 7.1). According to (7.5), the autocorrelation function  $R_u$  can be calculated as time average

$$R_u(\tau) = \lim_{T \rightarrow \infty} \left( \frac{1}{2T} \int_{-T}^T \Delta u(t) \Delta u(t + \tau) d\tau \right). \quad (7.11)$$

For a stationary random process  $R_u$  does not depend on time but only on  $\tau$ . Thus, the integration over  $t$  can be performed in (7.9) leading to

$$\langle |\Delta \underline{U}|^2 \rangle = 2T \int_{-2T}^{2T} R_u(\tau) \exp(-j\omega\tau) d\tau. \quad (7.12)$$

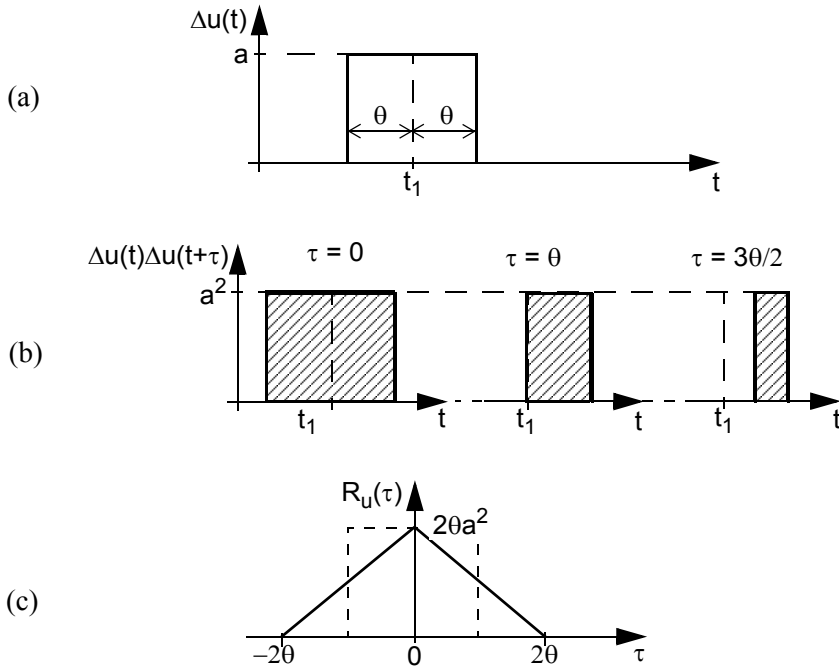


Fig. 7.1: Visualization of the autocorrelation function for a time dependent process: (a) single event of the noise signal  $\Delta u$  with amplitude  $a$ ; (b) integrand of (7.11) for different values of  $\tau$ ; (c) result after integration corresponding to the autocorrelation function.

Since for random processes  $R_u$  is typically non-zero only within a very small time interval of the duration of a characteristic time, the integration limits can be extended to infinity without changing the integral:

$$H_u(\omega) = \lim_{T \rightarrow \infty} \left( \frac{1}{T} \frac{\langle |\Delta \underline{U}|^2 \rangle}{2} \right) = \int_{-\infty}^{\infty} R_u(\tau) \exp(-j\omega\tau) d\tau \quad (7.13)$$

The new variable  $H_u$  is obviously the Fourier transform of the autocorrelation function and, according to the middle term, is directly related to the r.m.s. power dissipated by noise in a  $1\Omega$  resistor.

In theory,  $R_u$  is often derived from time domain considerations, e.g. particle movement due to collisions, which then allows to calculate  $H_u$  using (7.13). Experimentally, the opposite approach is usually taken in which  $H_u$  is measured and the inverse transform is applied to obtain

$$R_u(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H_u(\omega) \exp(j\omega\tau) d\omega, \quad (7.14)$$

allowing to draw conclusions about the physical nature of the measured fluctuations. The Fourier pair (7.13) and (7.14) is referred to as Wiener-Khinchine theorem.

Replacing in (7.10)  $\tau$  by  $-\tau$  and then  $t_1 - \tau$  by  $t_2$  reveals that  $R_u$  is an even function of  $\tau$ , i.e.  $R_u(-\tau) = R_u(\tau)$ . Hence from (7.13)

$$H_u(\omega) = 2 \int_0^{\infty} R_u(\tau) \cos(\omega\tau) d\tau. \quad (7.15)$$

With  $R_u$  even and replacing  $\tau$  by  $-\tau$  in (7.13) follows that also  $H_u$  is an even function of  $\omega$ , i.e.  $H_u(-\omega) = H_u(\omega)$ . Therefore, one can focus on just the positive frequency range and define the noise *power spectral density* (PSD) as

$$S_u(\omega) = 2H_u(\omega) = 4 \int_0^{\infty} R_u(\tau) \cos(\omega\tau) d\tau \text{ with } \omega \geq 0. \quad (7.16)$$

Furthermore, setting  $\tau = 0$  in (7.10) and  $\omega = 2\pi f$  gives with (7.14) and (7.16) a relation between the variance of the time domain signal and the PSD:

$$R_u(0) = \langle \Delta u(t)^2 \rangle = \int_{-\infty}^{\infty} H_u(f) df = \int_0^{\infty} S_u(f) df. \quad (7.17)$$

In most practical applications, noise is of interest in a *narrow* frequency band  $\Delta f$  around a center frequency  $f_0 \geq 0$ . In other words, the noise power of interest is concentrated within a narrow frequency range that can be defined by

$$\Delta f = \frac{1}{S_u(f_0)} \int_0^{\infty} S_u(f) df. \quad (7.18)$$

In this case, (7.17) yields for arbitrary  $f_0 \rightarrow f$

$$S_u(f)\Delta f = \langle \Delta u(t)^2 \rangle. \quad (7.19)$$

The narrow-band noise PSD corresponds to the variance of the time dependent noise signal. Also, the spectral density  $H_u$  corresponds in a 1 Hz frequency interval to the r.m.s. power dissipated in a 1  $\Omega$  resistor.

According to [3] the following two fundamental noise mechanisms are carrier scattering and trapping of carriers between, e.g., doping levels or traps to valence or conduction band. These fundamental mechanisms then lead to the, in the literature, often discussed “velocity” fluctuation (often also called diffusion noise) and “carrier number fluctuation” as well as “elementary event” based fluctuations [7] such as shot and flicker noise. An overview on fundamental calculation methods can be found in, e.g., [5, 7], but unfortunately on just a mathematical basis that lacks the physical insight.

### 7.1.2 Thermal noise

At a given lattice temperature  $T$  and under equilibrium conditions carriers in a semiconductor region, such as a resistor, possess thermal (i.e. kinetic) energy. Scattering with lattice atoms events leads to random motion. This random motion of carriers in metals was identified as the origin of thermal noise by Johnson in 1928 [8]. At the same time Nyquist provided a theoretical derivation of the thermal noise spectral density [9]. Its general formulation for the noise voltage PSD of a resistor  $R$  reads [9, 1]

$$S_V = 4 \left[ \frac{hf}{2} + \frac{hf}{\exp\left(\frac{hf}{k_B T}\right) - 1} \right] R \quad (7.20)$$

and includes so-called quantum noise (through the electromagnetic energy  $hf$ , with  $h$  as Planck constant). The first term in the brackets corresponds to “zero-point” energy and appears to be a contentious issue in the literature [1]. Figure 7.2 shows the normalized PSD as function of frequency for different practically interesting temperatures. At liquid He and N quantum

noise according to (7.20) has to be taken into account in high-frequency applications. However, at temperatures above 200 K as assumed in this book, the corresponding minimum frequency at which  $hf$  equals  $k_B T$  is 4.2 THz. Therefore, for practical applications  $hf \ll k_B T$  and (7.20) reduces to

$$S_V = 4k_B TR. \quad (7.21)$$

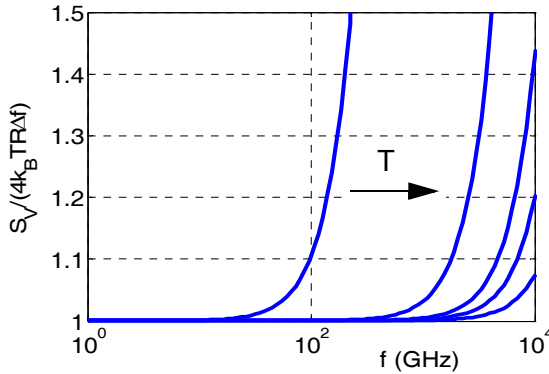


Fig. 7.2: Normalized PSD according to (7.20) vs. frequency with the temperature as parameter:  $T/K = 4.2, 77, 200, 300, 500$ .

Equation (7.21) is often referred to as Nyquist relation and generally being used for describing the noise voltage related PSD of resistive regions. Since  $S_V$  represents the variance of the voltage fluctuations at the terminals of the sample, the associated equivalent circuit, shown in Fig. 7.3, contains a noise voltage generator  $\sqrt{S_V}$  in series to the sample resistor  $R$ . According to the Norton-Thevenin theorem, this equivalent circuit can be converted to one with the conductance  $G = 1/R$  in parallel to a noise current source that has the noise current related PSD

$$S_I(f) \cong 4k_B TG. \quad (7.22)$$

The parallel equivalent circuit representation is more convenient for circuit *simulation* since it avoids creating undesired additional nodes.

Although originally derived for a metal resistor above expressions can also be applied to linear resistive regions in semiconductors since the mechanism is the same. However, in semiconductors other effects, such as



generation-recombination-trapping noise leading, e g., to carrier number fluctuations, can also play a significant role.

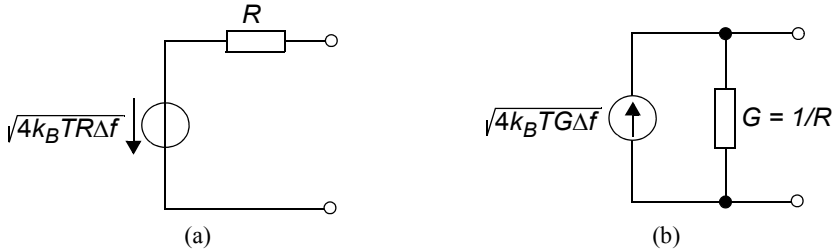


Fig. 7.3: Equivalent circuits for noise calculations at resistor terminals: (a) voltage fluctuation and (b) current fluctuation.

The discussion above assumed *equilibrium*, and the result does not depend on the current flowing through the probe. The question is how fluctuations can be described for a non-zero current. As long as the drift velocity is small compared to the thermal velocity, the carrier motion is still mainly determined by random motion except that the overall ensemble is moving in field direction. Often, the carrier ensemble is assumed to be in equilibrium with itself, and the carrier mean thermal energy is described by a carrier temperature  $T_c$ . This allows (7.22) to be written as

$$S_I(f) \cong 4k_B T_c G, \quad (7.23)$$

which is often also used to describe more general conditions (such as high fields). This equation needs to be modified if the number fluctuation becomes also relevant and the drift velocity approaches the thermal velocity.

Nyquist noise in form of (7.22) and (7.23) applies to *majority* currents since the conductance  $G$  (or resistance  $R$ ) is determined by the majority carrier density. For minority carriers, the more general expression for the variance of the current fluctuation,

$$\langle \Delta I(t)^2 \rangle \cong \left( \frac{q}{\Delta x} \right)^2 [\langle N \rangle \langle \Delta v_x^2 \rangle + \langle \Delta N^2 \rangle \langle v_x \rangle^2], \quad (7.24)$$

is applicable. The first term in the brackets contains the variance of the velocity fluctuation for a constant number of carriers and, hence, represents thermal noise. Note, that the velocity of minority carries generally is determined by drift *and* diffusion. The second term is proportional to the vari-

ance of the number fluctuation and will turn out to be shot or g-r-t noise (see later). Obviously, the variance of the current fluctuation (and respective noise PSD) generally consists of these two different types of noise.

For calculating thermal noise in nonlinear devices the small-signal conductance  $G = dI/dV$  in a given bias point has to be inserted in (7.23). In highly-doped degenerated regions the interaction between mobile carriers (i.e. carrier-carrier scattering and screening) may have to be taken into account through a corresponding correlation term.

### 7.1.3 Shot noise

Assume a space-charge region of width  $w$  with negligible generation or recombination, and with sufficient reverse bias to cause the drift velocity  $v_d$  to be much larger than the thermal velocity. Carriers crossing this SCR then have negligible velocity fluctuation and the first term in the brackets of (7.24) can be neglected.

All carriers are assumed to be independent from each other and have a characteristic time dependence  $qu(t-t_k)$  with  $t_k$  determining the random occurrence of the carrier in the zone  $\Delta x$ . Also, the occurrence is sharply concentrated in time around  $t_k$ . The stochastic process  $u$  in (7.11) must then be replaced by the sum of the single carrier events. Hence, averaging over  $T$  gives

$$R_I(\tau) = q^2 \lim_{T \rightarrow \infty} \left( \frac{\langle N(T) \rangle}{T} \int_{-T/2}^{T/2} u(t') u(t' + \tau) dt' \right). \quad (7.25)$$

The expression

$$\lim_{T \rightarrow \infty} \left( \frac{\langle N(T) \rangle}{T} \right) = \lambda \quad (7.26)$$

represents the number of occurring events per time, i.e. carriers passing through  $\Delta x$ . With the assumptions made earlier the exact shape of  $u$ , i.e. how the carrier crosses the zone boundaries, does not matter regarding the final result. Thus, one can simply set  $u(t-t_k) = \delta(t-t_k)$ . Also, for a stationary process the value of  $R_I$  does not depend on time, so one can set  $t' = 0$ . As a consequence,

$$R_I(\tau) = q^2 \lambda \int_{-\infty}^{\infty} \delta(0) \delta(0 + \tau) d\tau = \lambda q^2 \delta(0). \quad (7.27)$$

According to (7.16), the PSD is

$$S_I(\omega) = 2 \int_{-\infty}^{\infty} R_I(\tau) \exp(-j\omega\tau) d\tau = 2q^2 \lambda. \quad (7.28)$$

(Note the extension of the lower limit to  $-\infty$ .) Inserting the mean value of the current,  $\langle I \rangle = q\lambda$ , finally yields

$$S_I(\omega) = 2q \langle I \rangle, \quad (7.29)$$

which is the result found by Schottky for shot noise. This equation has been derived in the literature in many different ways (e.g. [1-4]).

#### 7.1.4 Recombination, generation, and trap noise

Noise from generation-recombination-trap (g-r-t) processes results in a fluctuation of the carrier number in a sample. The first fluctuation mechanism of interest is the SRH process in which carriers entering a region are randomly trapped in and released from trap levels. The resulting current through the region thus exhibits a random g-r-t related variation around a mean value  $I_{gr,SCR}$ . Depending on whether an SCR is forward or reverse biased, one obtains a slightly different PSD [1-5]

$$S_{gr} = k_{gr,SCR} 2q I_{gr}. \quad (7.30)$$

For the general case of a non-symmetrically doped junction,  $k_{gr,SCR} = 1$  at forward bias. At reverse bias also  $k_{gr,SCR} = 1$  for frequencies that are lower than the reciprocal of the trap occupation time, while  $k_{gr,SCR} = 2/3$  for higher frequencies. The derivation of these results is omitted here since bipolar transistors in practical noise sensitive applications will not be operated at such injection conditions at which SCR related g-r-t noise is relevant.

The second mechanism of interest in general is recombination noise in *neutral* regions. This can be caused by majority carrier fluctuations due to incomplete ionization and traps as well as by injection of minority carriers across a forward biased junction. Noise resulting from the former mechanism is negligible in Si above about 150K due to complete ionization. The

PSD corresponding to minority carrier injection can be calculated based on the combined continuity and transport equation and applying, e.g., the Langevin approach. Note though, that for bipolar transistors only very short neutral region widths are of interest, within which recombination is extremely small and usually negligible in practical applications. The case of minority carrier injection will be treated in detail later on for the neutral base region.

The third mechanism is impact ionization (II) noise resulting from random carrier pair generation by collisions within reverse biased SCRs. The impact ionization process itself has already been discussed in sec. 3.3.4. At this point the PSD resulting from the fluctuating carrier number caused by the random II and the practical case of weak avalanche is of interest. Focusing on the electron (transfer) current entering the BC SCR, the change of current in the SCR at a location  $x$  and within the interval  $dx$  is given according to (3.115) and (3.121) by

$$dI_n = MI_T\alpha_n dx. \quad (7.31)$$

The PSD of the current flowing through this interval  $dx$  corresponds to full shot noise. The overall noise results from multiplying each II event creating an electron (with charge  $-q$ ) by the factor  $M$ . Hence, replacing  $q$  by  $Mq$  in (7.25) and also in (7.28) yields

$$S_{dI_n} = (2qMI_T\alpha_n dx)M^2. \quad (7.32)$$

With the justified assumption that the (weak) II events within the SCR are uncorrelated, the total PSD can be obtained by integration:

$$S_C = 2qM^3I_T \int_0^{w_{BC}} \alpha_n dx + M^2(2qI_T). \quad (7.33)$$

The second term on the r.h.s. corresponds to the regular shot noise observed from the collector current leaving the SCR. Here again, each carrier entering the SCR generates in average  $M$  new carriers which leads to the factor  $M^2$  once the variance is calculated using (7.25) and (7.28). Using (3.119) the final result for the noise PSD then reads

$$S_C = 2qI_TM^3. \quad (7.34)$$

A more detailed consideration taking into account the difference between the electron and hole ionization rate, i.e. setting  $k_a = \alpha_p/\alpha_n$  gives for the case of electrons constituting the primary current entering the SCR [10]

$$S_C = 2qM^3 I_T \left[ 1 - (1 - k_a) \left( \frac{M-1}{M} \right)^2 \right] \quad (7.35)$$

with  $M$  as overall multiplication factor. In silicon,  $k_a \approx 0.3$  so that the noise is actually smaller than in the simple result (7.34).

### 7.1.5 Flicker or 1/f noise

Initially 1/f noise was believed to be a *surface* or *interface* effect. In an early theory [11] carriers at the semiconductor-oxide interface were assumed to tunnel into (and out of) traps with random time constants. Each trap level contributes a Lorentzian spectrum with its own characteristic time constant  $\tau_i$ , so that for a given current  $I$  the overall spectral density reads

$$S_I(f) = I^2 \sum_{i=1}^{\infty} \frac{K_i \tau_i}{1 + (\omega \tau_i)^2} \quad (7.36)$$

For a sufficiently wide range of time constants this superposition of independent processes gives a 1/f dependence. The term flicker noise comes from (older) observations of this low-frequency noise.

1/f noise has been the subject of many investigations, in which their origin is often assumed to be fluctuations in mobility or carrier number or both (e.g. [1, 2, 4, 12]. Mobility fluctuations were assumed to be caused by carrier scattering with slowly varying phonons [1], while carrier number fluctuations can result from g-r-t mechanisms not only at surfaces but also in *bulk* regions. Interpreting  $\tau_i$  in (7.36) as trap or collision time constants provides the necessary theoretical basis for describing both mobility and number fluctuation [3]. As another mechanism, lattice temperature variations, which can vary with frequencies in the order of the 1/f corner frequency, was suggested in [13].

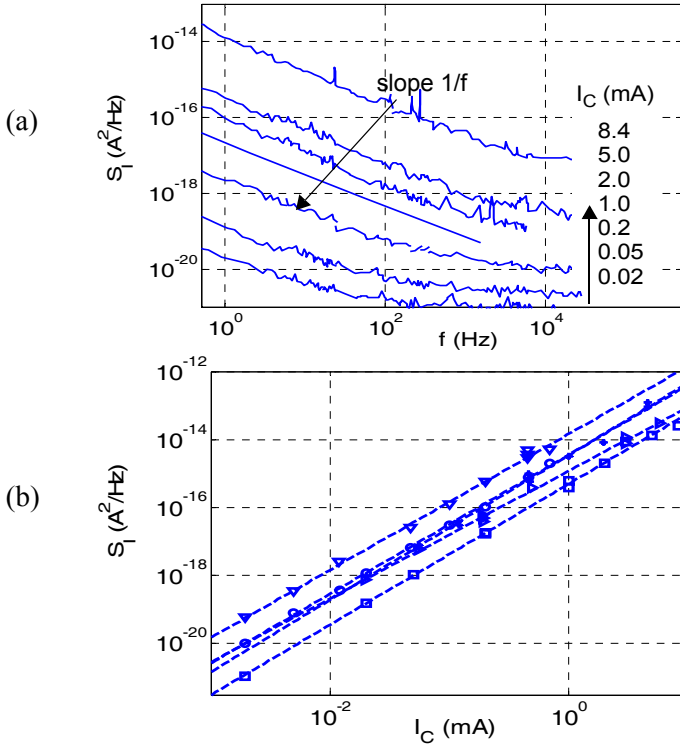


Fig. 7.4: Typical experimental low-frequency noise data in a large area BJT: (a) Collector current spectral density vs. frequency for collector currents varying from low-injection to about peak  $f_T$ . (b) Collector current spectral density vs. collector current at a fixed (low) frequency for different emitter window dimensions and configurations  $(0.4 \times (7, 14) \mu\text{m}^2)$  CEB and  $(0.4, 0.8, 1.2) \times 14 \mu\text{m}^2$  CBEBC).

For device modeling, it is more convenient to formulate (7.36) in terms of the current flowing through the sample. Compiling many results obtained for semiconductor devices, the current spectral density of flicker noise is generally described by

$$S_I = K_F \langle I \rangle^{a_F} / f^b \quad (7.37)$$

with  $K_F$  as prefactor depending on geometry and temperature.  $1/f$  noise is obtained for  $b = 1$ , while  $b = 2$  is generally associated with burst noise.

According to [14],  $a_F = 1$  corresponds to (bulk) generation-recombination noise in the SCR of a non-linear resistance, such as a forward biased

junction. In contrast,  $a_F = 2$  indicates g-r-t noise, i.e. carrier number fluctuation, in a linear resistance. Figure 7.4a shows the behavior of the collector current related spectral density vs. frequency for different currents of a BJT. A  $1/f$  dependence is observed by comparison to the dashed reference line. Figure 7.4b exhibits the current dependence of  $S_{I_C}$  at a low frequency. The results clearly follow a  $I_C^2$  dependence. These measurements also did not show a dependence on  $V_{BC}$ , indicating that the  $1/f$  noise does not seem to originate from bulk generation-recombination. SiGe HBTs show a very similar behavior according to [15].

## 7.2 Intrinsic transistor

### 7.2.1 Neutral base region

Transport in the neutral base region is determined by minority carriers. Hence, the questions here are how can minority carrier noise in neutral regions be described and what is its frequency dependence. Similar to classical theory, the minority continuity and transport equation need to be solved with properly added spatially distributed fluctuation sources. Below, such a solution will be performed for the neutral base region of an npn transistor based on the following assumptions:

- (i) 1D case;
- (ii) neglecting g-r-t noise, which is very small in advanced transistors;
- (iii) neglecting the drift field; i.e. considering a diffusion transistor.

The purpose of the last assumption is to simplify the already very lengthy derivation while showing the basic principle and maintaining the fundamental results. From those an extension to the drift field case will then be possible without going through the solution process again. Dropping the Poisson equation corresponds implicitly to the assumption of negligible recombination related fluctuations resulting from doping levels. With above assumptions, only the electron continuity equation (2.8b) needs to be considered, which then reads

$$\frac{dJ_{nx}}{dx} - q \frac{\partial n}{\partial t} = 0. \quad (7.38)$$

Dropping the field term in (2.9b) yields the electron transport equation

$$J_{nx} = qD_{nB} \frac{dn}{dx} \quad (7.39)$$

with  $D_{nB} = \mu_{nB}V_T$  as spatial average of the diffusion constant.

Fluctuations can be described by introducing random sources at locations  $x'$  within the neutral base and by observing their impact on the current density at arbitrary locations  $x$  and, in particular, at the ends of the neutral base. In the case considered here, the g-r-t mechanism is neglected in the continuity equation. However, a random current related fluctuation term  $\delta J_{nx}$  is added to the r.h.s. of the continuity equation (7.38), leading to a small-signal perturbation of the carrier density. The PSD of  $\delta J_{nx}$  is given by (7.22) with  $G = q\mu_{nB}n(x')/w_B = q^2D_{nB}n(x')/(k_BT w_B)$  and reads

$$S_{\delta j_n}(x'_1 - x'_2) \cong 4q^2 \frac{D_{nB}n(x')}{w_B} \delta(x'_1 - x'_2) \quad (7.40)$$

for a given bias point with  $n(x')$ . Summing (i.e. integrating) over the PSD of all independent current fluctuations (i.e. noise sources) *distributed* within the neutral base then gives the PSD at a selected location  $x$ .

Since noise is a small signal process, the carrier and current density are expressed as the sum of a DC value and a superimposed AC signal. This allows to separate above equations into a DC and an AC portion. Linearization and transformation into frequency domain gives after inserting  $\delta j_n(x, \omega)$  into (7.38) the stochastic (small-signal) continuity equation,

$$\frac{d\delta J_{nx}}{dx} - j\omega q \delta \underline{n} = \frac{d\delta j_n}{dx}, \quad (7.41)$$

with  $\delta \underline{n}$  and  $\delta \underline{J}_{nx}$  as stochastic harmonic small-signal components. This method is often referred to as Langevin approach (e.g. [1, 2, 4, 5, 7]). According to [7], the corresponding Langevin sources must be of white noise type. According to [3] this assumption is not required. An alternative method for calculating the PSD of the current entering and leaving the neutral base is based on using Carson's theorem (e.g. [2]).

Finally, with

$$\delta \underline{J}_{nx} = qD_{nB} \frac{d\delta \underline{n}}{dx} \quad (7.42)$$

one obtains the stochastic diffusion equation for the neutral base,



$$qD_{nB} \frac{d^2 \delta \underline{n}}{dx^2} - j\omega q \delta \underline{n} = \frac{d\delta j_n}{dx}, \quad (7.43)$$

which describes the behavior of the carrier density fluctuation with respect to excitations due to (internal) noise sources. This equation will be solved below. Then, the spatially dependent current density fluctuation will be calculated from which the PSD of a current at a selected location  $x$  in the neutral base can be derived.

Using the normalized variables  $x/w_B \rightarrow x$  and  $x'/w_B \rightarrow x'$  as well as defining  $\underline{k}^2 = j\omega w_B^2/D_{nB}$  results in

$$\frac{d^2 \delta \underline{n}}{dx^2} - \underline{k}^2 \delta \underline{n} = \frac{w_B}{qD_{nB}} \frac{d\delta j_n}{dx}. \quad (7.44)$$

The electron density at  $x = 0$  is determined by the applied voltage and, thus does not fluctuate. At the collector end,  $x = 1$ , an infinite carrier velocity is assumed. Therefore, the boundary conditions are simply

$$\delta \underline{n}(x = 0) = \delta \underline{n}(x = 1) = 0. \quad (7.45)$$

This corresponds to short circuit conditions. Since the homogeneous solution of (7.44) is the trivial solution and the inhomogeneous term is given by a delta function, the Green's function technique [16, 17] is usually preferred for solving (7.44). The Green's function determines the impact of a (current) fluctuation at location  $x'$  on the carrier density fluctuation at a selected location  $x$ . According to [16, p. 509], the Green's function for the homogeneous part of (7.44) is

$$\underline{G}(x, x', \omega) = \begin{cases} -\frac{\sinh(\underline{k}x) \sinh(\underline{k}[1-x'])}{\underline{k} \sinh(\underline{k})} & , \quad 0 \leq x \leq x' \leq 1 \\ -\frac{\sinh(\underline{k}x') \sinh(\underline{k}[1-x])}{\underline{k} \sinh(\underline{k})} & , \quad 0 \leq x' \leq x \leq 1 \end{cases}. \quad (7.46)$$

The general solution then reads

$$\delta \underline{n}(x, \omega) = \frac{w_B}{qD_{nB}} \int_0^1 \underline{G}(x, x', \omega) \frac{\partial \delta j_n}{\partial x'} dx'. \quad (7.47)$$

The noise induced minority current fluctuation is then obtained by inserting (7.47) into the transport equation (7.42) (with normalized variable  $x$ )

$$\delta J_{nx}(x, \omega) = \int_0^1 \underline{G}_x(x, x', \omega) \frac{\partial \delta j_n}{\partial x'} dx' \quad (7.48)$$

with the derivative of the Green's function w.r.t.  $x$

$$\underline{G}_x(x, x', \omega) = \frac{\partial \underline{G}(x, x', \omega)}{\partial x}. \quad (7.49)$$

The PSD of  $\delta J_{nx}(x, \omega)$  requires the PSD of the injected fluctuation  $\delta j_n$ . The derivative term of  $\delta j_n$  in (7.48) is eliminated by partial integration resulting in

$$\int_0^1 \underline{G}_x \frac{\partial \delta j_n}{\partial x'} dx' = (\underline{G}_x(x, x', \omega) \delta j_n) \Big|_0^1 - \int_0^1 \underline{G}_{xx'} \delta j_n dx'$$

The first integral on the r.h.s. is zero, since  $G_x = 0$  for  $x' = 0 \leq x$  and for  $x \leq x' = 1$ . Hence, the current density fluctuation is given by

$$\delta J_{nx}(x, \omega) = - \int_0^1 G_{xx'} \delta j_n dx' \quad (7.50)$$

with

$$\underline{G}_{xx'} = \underline{k} \begin{cases} \frac{\cosh(\underline{k}x) \cosh(\underline{k}[1-x'])}{\sinh(\underline{k})} & , 0 \leq x \leq x' \leq 1 \\ \frac{\cosh(\underline{k}x') \cosh(\underline{k}[1-x])}{\sinh(\underline{k})} & , 0 \leq x' \leq x \leq 1 \end{cases}. \quad (7.51)$$

The PSD of  $J_{nx}$  is the sum of the PSDs of all independently induced fluctuations over the neutral base region. According to (7.7), (7.13) and (7.16) the corresponding mathematical formulation involves the product of  $\delta J_{nx}(x, \omega)$  with its conjugate complex. Taking this at two different locations  $x_1$  and  $x_2$  allows to write a general expression that can also be used for calculating the correlation between the 1D terminal currents:

$$S_{J_{nx}}(x_1, x_2, \omega) = \int_0^1 \int_0^1 S_{\delta j_n}(x'_1 - x'_2) \underline{G}_{xx'}(x_1, x'_1, \omega) \underline{G}_{xx'}^*(x_2, x'_2, \omega) dx'_1 dx'_2. \quad (7.52)$$

For evaluating the integral (7.40) is inserted. Furthermore, according to the transport equation (7.39), the DC carrier distribution of the *diffusion* transistor,  $n(x') = n(0)(1 - x')$ , is used here.

It is given by (3.45) with  $\zeta = 0$  and  $n_c = 0$  which also leads to

$$n(0) = \frac{I_T w_B}{q A_E D_{nB}}. \quad (7.53)$$

Inserting above expressions into (7.52) finally gives

$$S_{J_{nx}}(x_1, x_2, \omega) = \frac{4qI_T}{A_E} \int_0^1 (1 - x') G_{xx'}(x_1, x', \omega) \underline{G}_{xx'}^*(x_2, x', \omega) dx'. \quad (7.54)$$

The transfer current PSD is then obtained by taking  $S_{J_{nx}}$  at  $x_1 = x_2 = 1$

$$S_{J_T}(\omega) = \frac{4qI_T}{A_E} \int_0^1 (1 - x') |G_{xx'}(1, x', \omega)|^2 dx' \quad (7.55)$$

with the lower part of (7.51),

$$G_{xx'}(1, x', \omega) = \underline{k} \frac{\cosh(\underline{k}x')}{\sinh(\underline{k})}. \quad (7.56)$$

The evaluation of the integral is quite lengthy and leads after multiplying with the emitter area to

$$S_{I_T}(\omega) = 2qI_T. \quad (7.57)$$

The transfer current noise PSD corresponds to full shot noise.

The PSD of the injected electron current density component  $J_{nE}$  at  $x = 0$  is given by  $S_{J_{nx}}$  in (7.54) at  $x_1 = x_2 = 0$ ,

$$S_{J_{nE}}(\omega) = \frac{4qI_T}{A_E} \int_0^1 (1 - x') |G_{xx'}(0, x', \omega)|^2 dx', \quad (7.58)$$

with the upper part of (7.51),

$$G_{xx'}(0, x', \omega) = \underline{k} \frac{\cosh(\underline{k}[1 - x'])}{\sinh(\underline{k})}. \quad (7.59)$$

Again, a lengthy evaluation yields

$$S_{I_{nE}}(\omega) = 2qI_T \left[ 2 \operatorname{Re} \left\{ \frac{y_{me}(\omega)}{g_{me}} \right\} - 1 \right], \quad (7.60)$$

where  $g_{me} = y_{me}(0)$  and

$$y_{me}(\omega) = \frac{I_{nE}}{V_{B'E}} = g_{me} \frac{k}{\tanh(k)} \quad (7.61)$$

is the input admittance in common-base configuration obtained from (4.190) for zero drift field. It can be shown though (by an even more lengthy derivation, e.g. [4]) that (7.57) remains valid also if a non-zero drift field as well as recombination are included. This also holds for (7.60) if the DC current  $I_T$  is just replaced by  $I_{nE} = I_E - I_{BE}$ . The frequency dependence of  $S_{I_{nE}}$ , normalized to the transfer current shot noise  $S_{I_T}$ , is shown in Fig. 7.5a. The normalized frequency value of 1 corresponds to the neutral base transit frequency (in a diffusion transistor). At this frequency  $S_{I_{nE}}$  is only slightly larger than  $S_{I_T}$ , but increases rapidly towards higher frequencies.

The base current flowing into the neutral base is  $i_{Bb} = i_{nE} - i_T$ , so that the corresponding time dependent fluctuation is  $\delta i_{Bb} = \delta i_{nE} - \delta i_T$ . This gives the corresponding noise PSD

$$S_{I_{Bb}} = S_{I_{nE}} + S_{I_T} - 2Re\{S_{I_{nE}I_T}\}. \quad (7.62)$$

Both  $S_{I_{nE}}$  and  $S_{I_T}$  have already been calculated in (7.57) and (7.60). The cross-correlation PSD is obtained from (7.54) as  $S_{J_{nE}J_T} = S_{J_{nx}}(1, 0, \omega)$  leading to

$$S_{J_{nE}J_T}(\omega) = \frac{4qI_T}{A_E} \int_0^1 (1-x') G_{xx'}(1, x', \omega) G_{xx'}^*(0, x', \omega) dx'. \quad (7.63)$$

Evaluation gives

$$2Re\{S_{I_{nE}I_T}(\omega)\} = 4qI_T Re\left\{\frac{y_m(\omega)}{g_m}\right\} \quad (7.64)$$

with  $g_m = y_m(0)$  and  $y_m$  as frequency dependent small-signal *transconductance* in common-emitter configuration. Inserting into (7.62) gives

$$S_{I_{Bb}} = 4qI_T \left[ Re\left\{\frac{y_{me}(\omega)}{g_{me}}\right\} - Re\left\{\frac{y_m(\omega)}{g_m}\right\} \right], \quad (7.65)$$

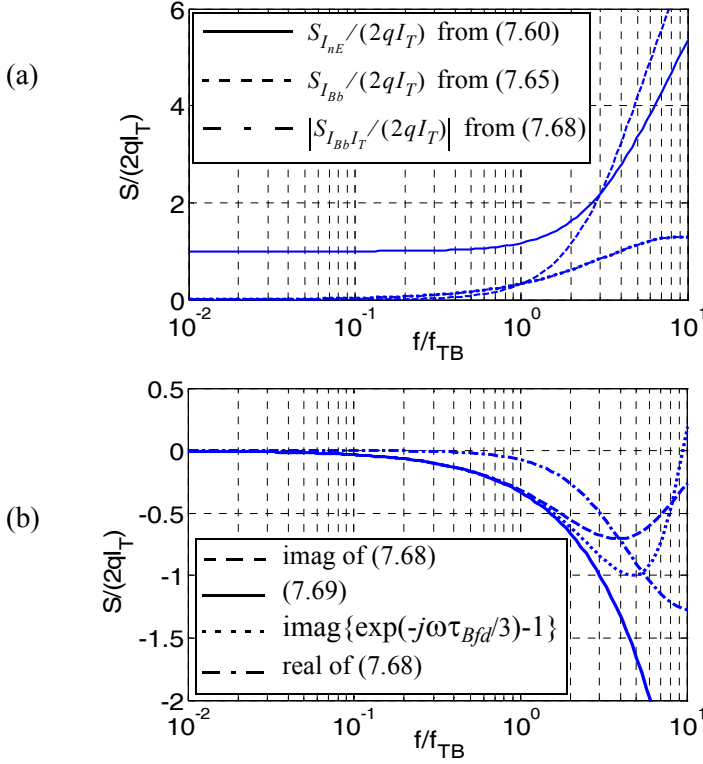


Fig. 7.5: Base region related normalized PSDs as function of normalized frequency with  $f_{TB} = (2\pi\tau_{Bfd})^{-1}$  as base region transit frequency. (a) Injection and base current PSD along with the magnitude of the correlation PSD for comparison. (b)  $\text{Im}\{S_{I_{Bb}I_T}\}/(2qI_T)$  along with  $-\omega\tau_{Bfd}/3$  from (7.69), the often found exponential approximation, and  $\text{Re}\{S_{I_{Bb}I_T}\}/(2qI_T)$  for comparison.

which approaches zero towards  $\omega = 0$  as expected since this is the PSD of the *dynamic* base current  $I_{Bb} = j\omega Q_{Bf}$  supporting the neutral base region only (since recombination can be neglected). The corresponding frequency behavior is shown in Fig. 7.5a. At about three times the base region transit frequency  $S_{I_{Bb}}$  starts to exceed  $S_{I_{nE}}$ . For not too high frequencies, series expansion of (7.65) yields the approximation

$$S_{I_{Bb}} \cong 2qI_T \left[ 8 \frac{(\omega\tau_{Bfd0})^2}{15} - \frac{(\omega\tau_{Bfd0})^2}{30} \right] = 2qI_T \frac{(\omega\tau_{Bfd0})^2}{2}. \quad (7.66)$$

The cross-spectral density representing the correlation between the *dynamic* base current  $I_{Bb}$  and the transfer current follows from

$$\underline{S}_{I_{Bb}I_T}(\omega) = S_{I_{nE}I_T}(\omega) - S_{I_T}(\omega). \quad (7.67)$$

Inserting (7.57) and  $S_{I_{nE}I_T}$  from (7.64) yields the correlation PSD

$$\underline{S}_{I_{Bb}I_T}(\omega) = 2qI_T \left( \frac{\underline{y}_m(\omega)}{g_m} - 1 \right). \quad (7.68)$$

The corresponding frequency behavior of the magnitude is shown in Fig. 7.5a for comparison. It increases slightly more rapidly at frequencies below  $f_{TB}$  than  $S_{IBb}$  but its peak at high frequencies is much lower than that of  $S_{IBb}$  and  $S_{InE}$ . Series expansion of  $\underline{y}_m(\omega)$  and truncating after the first frequency term leads for the diffusion transistor considered here to

$$S_{I_{Bb}I_T}(\omega) \cong -j2qI_T \frac{\omega \tau_{Bfd}}{3}. \quad (7.69)$$

From this result the following conclusions can be drawn. First, correlation *reduces* noise at *high* frequencies. Second, it is closely related to the NQS effect in the transfer current since its frequency dependence is determined by the NQS expression and time constant of the transconductance. Third, for the same reason as for the NQS solution there is no delay term of the form  $\exp(-j\omega\tau_{Bfd})$ , which is often found though in the literature (e.g. [15,18]). However, as Fig. 7.5b shows, at frequencies beyond  $f_{TB}$  the exponential term (by coincidence in fact) turns out to be a somewhat more accurate approximation than (7.69). Interpreting  $I_T\tau_{Bfd}$  as neutral base minority charge yields for (7.69) the same expression as in [19] in which charge partitioning is employed for modeling noise correlation. However,  $I_T\tau_{Bfd}$  only corresponds to the neutral base charge at *low injection*.

The impact of a base drift field requires to include the drift term in the transport equation. This then leads to a different Green's function solution instead of (7.46). As an approximation for practical purposes, just the corresponding transit time and NQS expression for  $\underline{y}_m(\omega)$  can be used in the above solution. Finally, note that fluctuations of majority carriers correspond to thermal noise and are negligible [20] at the injection levels considered in the solutions given above.

### 7.2.2 Base-collector space-charge region

Of most concern for practical applications is the noise in the collector current. Its largest contribution generally stems from the noise in the transfer current. A simplistic view is to just consider the fluctuations of the current leaving the BC SCR, which corresponds to shot noise and a majority carrier point of view. As a consequence, the PSD of the transfer current is according to (7.29)

$$S_{I_T}(\omega) = 2qI_T \quad (7.70)$$

with  $I_T$  as the DC transfer current.

If in a compact model the transfer current is defined by the current leaving the BC SCR, the correlation between this current and the (dynamic) base current needs to be modified with respect to the delay of the carriers drifting through the BC SCR. In this case, it is feasible to use an ideal delay term so that the modified form of (7.68) reads

$$S_{I_{Bb}, I_T}(\omega) = 2qI_T \left( \frac{y_m(\omega)}{g_m} e^{-j\omega\tau_{BC}} - 1 \right). \quad (7.71)$$

Series expansion of the exponential term and maintaining only the first-order frequency expression leads for the diffusion transistor considered here to

$$S_{I_{Bb}, I_T}(\omega) \cong -j2qI_T\omega \left( \frac{\tau_{Bfd}}{3} + \tau_{BC} \right). \quad (7.72)$$

This corresponds to an addition of the delay times of each region. Using the exponential approximation  $[\exp(-j\omega\tau_{Bfd})-1]$  for  $y_m/g_m$  increases the accuracy w.r.t (7.71) compared to (7.72). This is true the larger the ratio  $\tau_{BC}/\tau_{Bfd}$  and the larger the base drift field is since pure drift regions correspond to an ideal delay, which is described best by  $\exp(-j\omega\tau_{BC})$ .

### 7.2.3 Neutral emitter

The backinjection of holes into the emitter leads generally to diffusion and recombination noise in both the mono-silicon and poly-silicon region. Advanced transistors have very shallow emitter junction depths and, thus, are transparent w.r.t. to the minority carrier transport. As a consequence,

only diffusion noise needs to be considered in the mono-silicon region. The corresponding PSD can be calculated directly based on the solution for the electron current injected into the base. Modifying (7.60) correspondingly yields

$$S_{I_{pB}}(\omega) = 2qI_{jBEi} \left[ 2 \operatorname{Re} \left\{ \frac{y_{jBEi}(\omega)}{g_{jBEi}} \right\} - 1 \right], \quad (7.73)$$

where  $g_{jBEi} = y_{jBEi}(0)$  and

$$y_{jBEi}(\omega) = \frac{I_{jBEi}}{V_{B'E'}} = g_{jBEi} \frac{k_{pE}}{\tanh(k_{pE})}. \quad (7.74)$$

with  $k_{pE} = \sqrt{2j\omega\tau_{pE}}$  and  $\tau_{pE}$  as emitter storage time.

Series expansion and truncation after the second-order term gives the result for not too high frequencies

$$S_{I_{pB}}(\omega) \cong 2qI_{jBEi} \left[ 1 + 16 \frac{(\omega\tau_{pE})^2}{15} \right]. \quad (7.75)$$

This corresponds to full shot noise at low frequencies. According to [20] fluctuations of majority carriers are negligible.

## 7.2.4 Impact ionization noise

At this point the question might be asked why II noise is relevant for bipolar transistors. Device operation at impact ionization bias conditions is usually tried to be avoided in circuit design, since the output conductance increases. On the other hand, the speed of high-performance SiGe HBTs generally increases with  $V_{CE}$ . Furthermore, the trend of SiGe HBT process development clearly shows an improvement in speed at the expense of a lower breakdown voltage and, thus, a reduced upper limit for the collector operating voltage. Therefore, for advanced technologies and applications, circuit optimization requires a careful compromise that may lead to device operation in weak avalanche operation. Such an optimization is only possible with a compact model that includes a description of II noise under avalanche conditions.

The result obtained in (7.35) was derived for strong impact ionization in a pn junction and includes the noise contribution of both the avalanche cur-



rent at the end of the SCR and the original (or primary) current entering the SCR. After rearranging, (7.35) can also be written as

$$S_C = 2qMI_T[2M - 1 + k(M - 1)^2]. \quad (7.76)$$

Subtracting the noise contribution of the primary current,  $2qM^2I_T$ , yields the PSD contribution of the avalanche current itself

$$S_{I_{AVL}} = 2qMI_T(M - 1)[1 + k(M - 1)]. \quad (7.77)$$

In most bipolar transistors and materials  $k$  is small. Also, only the weak avalanche regime is of interest. Remembering (3.122) allows to express the approximated impact ionization related PSD by the avalanche current:

$$S_{I_{AVL}} \cong 2qMI_{AVL}. \quad (7.78)$$

This simple compact description has demonstrated to be sufficiently accurate for practical applications [21, 22].

## 7.3 Multi-dimensional effects

### 7.3.1 Internal transistor

As described in sec. 5.2 the lateral extension of the internal transistor to a 2D structure can lead to emitter current crowding. The impact of the related distributed effects and, in particular, of the internal base resistance  $R_{Bi}$  on noise behavior of the transistor was investigated in, e.g., [23-26] for both DC and small-signal dynamic (AC) current crowding. While in [23-25] a multi-transistor approach was used in [26] analytical formulations were sought that are suitable for compact modeling.

For noise-sensitive practical applications such as low-noise amplifiers, transistors with relatively long rectangular emitter stripes are used both to reduce the noise contribution of the base resistance and to keep the input impedance close to 50  $\Omega$ . According to the literature given above there is no suitable compact solution for the noise contribution from the internal base resistance as function of DC current crowding. However, as it was shown in sec. 5.2, DC current crowding is negligible in modern HBT and BJT technologies due to the achievable narrow emitter widths which are

preferred in practical noise-sensitive applications. As a consequence, the contribution of  $R_{Bi}$  can be represented by thermal noise

$$S_{I_{R_{Bi}}} = 4k_B T / R_{Bi}, \quad (7.79)$$

in which  $R_{Bi}$  is bias (and geometry) dependent as described in sec. 5.2.

At high frequencies AC emitter current crowding may become significant. While this changes the transfer function for the  $R_{Bi}$  related noise impact on the output node it does not change the PSD that needs to be assigned to the compact element  $R_{Bi}$  given in (7.79). Any frequency dependent addition to (7.79) would be of second-order. As shown in [26] there is also no correlation between the PSD of  $R_{Bi}$  on one hand and the PSDs of either the intrinsic base or the transfer current (extended to the internal transistor region) on the other hand. As a consequence, the lumped representation consisting of  $R_{Bi}$ ,  $I_{Bi}$ , and  $I_T$  is a valid and accurate description of noise for the internal transistor.

### 7.3.2 External (access) regions

The nodes C', B\*, and E' of the internal transistor are connected to the terminal nodes by ohmic regions, which generally can be described by bias independent series resistances. The corresponding fluctuations are of thermal nature so that (7.22) or (7.21) do apply. This would still be the case, if possible distributed effects at high frequencies are taken into account using a similar approach as for the internal base resistance. Again, the possible frequency dependent AC current crowding function only alters the transfer function, but the resistor noise remains thermal (to first order).

### 7.3.3 Flicker and 1/f noise

In a BJT, there are a number of possibly relevant locations for FN. First, traps in space-charge regions, at surfaces, at the poly-mono-silicon interface of the emitter, at the poly-Si grain boundaries, and in the (bulk) base region all cause g-r-t type noise. Second, self-heating induced temperature variations may add to 1/f noise. In HBTs, a larger trap density may exist due to material composition related lattice strain, especially for very sharp compositional transitions in the BE and BC SCR. Moreover, the more ad-

vanced the process is the shorter are the annealing cycles and, hence, the higher the probability becomes for an increase in trap density. These FN related mechanisms and their relative impact shall be discussed briefly below, based on the results obtained in the literature. In contrast, g-r-t noise from doping atom ionization in neutral regions can be safely neglected in BJTs and HBTs for temperatures above 200K due to the high doping concentrations in these regions (e.g. [4]).

The actual origin of FN sources in BJTs and HBTs has been the topic of many investigations and papers. This is mostly due to the fact that this noise type is process and sometimes even device type dependent, making it difficult to draw general conclusions for a particular technology (such as SiGe HBTs or III/V HBTs). In other words, FN and  $1/f$  noise are difficult to predict and always have to be measured for each process technology.

In early BJT generations, FN was assumed to be caused by surface states in the extrinsic base region and by g-r-t traps in the SCRs. Experimental investigations of transistors with  $\sim 0.5\mu\text{m}$  base width and  $\sim 5\mu\text{m}$  emitter width in the late seventies [27] found the  $1/f$  noise spectral density

$$S_{IB}(f) = K \frac{A_{E0}^{3/2} J_B^{a_F}}{w_B f} \quad (7.80)$$

to be dependent only on the DC base current density  $J_B$ , the emitter window area  $A_{E0}$ , and the base width  $w_B$ . The current exponent factor  $a_F = 2m_{BE}/m_{BER}$  equals 1 in the medium bias region and 2 in the low-injection (i.e. recombination) bias region. Above equation results from assuming base sheet resistance fluctuations as  $1/f$  noise origin, since surface related noise turned out to be a negligible contributor. Fluctuations in base sheet resistance can be caused by mobility, carrier number, and g-r-t (via SCR width) fluctuations.

The introduction of poly-silicon emitters led to additional traps located within the interfacial oxide and the poly-silicon grain boundaries. Although the FN spectral density can vary significantly with the interface treatment during processing [28], most experimental investigations of such processes indicate the interfacial oxide to be the main contributor to FN (e.g. [29-33]). Typical data taken in the medium injection range (i.e. at constant current gain), as shown in Fig. 7.4b, clearly exhibit  $a_F = 2$ , which

indicates a g-r-t mechanism as noise origin. Writing (7.37) or (7.80) in the form

$$S_{IB}(f) = K_F \frac{I_B^2}{f} \quad (7.81)$$

allows to determine  $K_F$  at a given current and frequency. According to Fig. 7.6,  $K_F$  is inversely proportional to  $A_{E0}$  but does not show a clear dependence on the emitter window perimeter  $P_E$ . Since  $K_F I_B^2 \sim A_{E0}$  this is a strong indication that  $1/f$  noise is proportional to the number of interface traps within the emitter window area. This result was found to be independent on  $V_{BC}$  even for a forward biased junction [31].

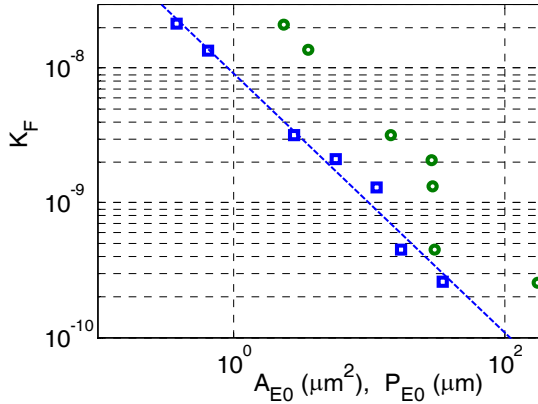


Fig. 7.6: Example for experimental dependence of flicker noise factor  $K_F$  on emitter window area  $A_{E0}$  (squares) and perimeter length  $P_{E0}$  (circles) for a 15 GHz BJT technology [31].

Existing SiGe HBTs with poly-silicon emitter do not exhibit fundamentally different behavior in terms of  $1/f$  noise compared to their BJT counterparts (e.g. [32, 33]). The reason for this is the dominance of the interfacial oxide related noise contribution on one hand and the apparently relatively small increase of (volume) trap density on the other hand. The latter is certainly, at least in part, a result of the graded material compositions for both Ge and C in real SiGe HBTs. The experimental results for different SiGe process generations in [32] seem to confirm the base width dependence of  $S_{IB}(f)$  in (7.80): for the same area, frequency and current

density an increase of  $S_{IB}(f)$  is observed in subsequent process generations with reduced base width. From the literature it is still unclear whether the Si/SiGe material transition induces additional traps and, thus, increases  $1/f$  noise. Since the current  $I_{BC}$  is much smaller than  $I_{BE}$ , and g-r-t noise related to the BE junction is already relatively small, any possibly existing g-r-t noise related to the BC junction is expected to be negligible for practical applications. This also holds for high current densities, when some of the possibly existing traps located in the region of the sharp Ge drop become exposed to the neutral base region, since this operating region is not useful anymore in circuits.

Following the previously mentioned notion, the total number of traps should decrease as  $A_{E0}$  shrinks. Then, according to (7.36), for large area transistors ( $n$  large) overall  $1/f$  behavior results, while for small transistors ( $n$  small) the different contributions of the Lorentzian spectra should be observed. This theory is indeed supported by the experimental results in, e.g., [31, 32]. Furthermore, different transistors with the same window area each exhibit a different spectral density. However, after superposition  $S_{IB}(f)$  approaches the  $1/f$  behavior and the magnitude expected from measurements on larger devices. Usually, even in today's most advanced devices, often such homogenous  $1/f$  behavior is observed, making (7.36) impractical for compact modeling. Hence, adding "lumped"  $1/f$  noise term to (7.36) yields a more flexible model expression

$$S_{IB}(f) = I_B^{a_F} \sum_{i=1}^n \frac{K_i \tau_i}{1 + (\omega \tau_i)^2} + K_F \frac{I_B^{a_F}}{f}. \quad (7.82)$$

The dependence of  $S_{IB}(f)$  on interfacial oxide thickness  $t_{ox}$  was investigated in [33]. Applying (7.81), it was found that  $K_F \sim \exp(t_{ox})$ . This is another strong indication that the interfacial oxide is the main origin for  $1/f$  noise.

## 7.4 References

- [1] A. Van der Ziel, "Noise in solid state devices and circuits", Wiley, NY, 1986.

- [2] M. Buckingham, "Noise in electronic devices and systems", Ellis Horward Ltd, Chichester, 1983.
- [3] C.-T. Sah, "Fundamentals of Solid-State Electronics", World Scientific, Singapore, 1991.
- [4] A. Blum, "Electronic noise", (in German) Teubner Stuttgart, 1995.
- [5] F. Bonani and G. Ghione, "Noise in semiconductor devices", Springer-Verlag Berlin Heidelberg, 2001.
- [6] A. Papoulis and S. U. Pillai, "Probability, Random Variables and Stochastic Processes", 4th Edition, Tata McGraw-Hill, 2002.
- [7] C. Van Vliet, "Macroscopic and microscopic methods for noise in devices", IEEE Trans. on Electron Dev., Vol. 41, No. 11, pp. 1902-1915, 1994.
- [8] J. Johnson, "Thermal agitation of electricity in semiconductors", Phys. Rev., Vol. 32, pp. 97-109, 1928.
- [9] W. Nyquist, "Thermal agitation of electric charge in conductors", Phys. Rev., Vol. 32, pp. 110-113, 1928.
- [10] R. McIntyre, "Multiplication noise in uniform avalanche diodes", IEEE Trans. Electron Dev., Vol. 13, No. 1, pp. 164-168, 1966.
- [11] A. McWerther, "1/f noise and Germanium surface properties", In Semiconductor Surface Physics, Univ. Pennsylvania Press, Philadelphia, 1957.
- [12] F. Hooge, "1/f noise sources", IEEE Trans. on Electron Dev., Vol. 41, No. 11, pp. 1926-1935, 1994.
- [13] L. Forbes, C. Zhang, B. Zhang, "Experimental verification of the dependence of bipolar transistor flicker noise on power dissipation", IEEE Trans. on Electron Dev., Vol. 49, No. 5, pp. 945-947, 2002.
- [14] T. Kleinpenning, "Location of low-frequency noise sources in sub-micrometer bipolar transistors", IEEE. Trans. Electron Dev., Vol. 39, pp. 1501-1506, 1992.
- [15] G. Niu, "Noise in SiGe HBT RF technology: Physics, modeling, and circuit implications", Proc. IEEE, pp. 1583-1597, 2005.
- [16] I. Bronstein and K. Semendjajew, "Taschenbuch der Mathematik", 20. Edition, Teubner, Leipzig, 1981.
- [17] Morse and Feshbach, Methods of Theoretical Physics, McGraw-Hill, 1978.
- [18] M. Rudolph, F. Korndörfer, P. Heymann, W. Heinrich, "Compact large-signal shot noise model for HBTs", IEEE Trans. Microwave Theory and Techn., Vol. 56, pp. 7-14, 2008.

- [19] J. Paasschens, R. Havens, L. Tiemeijer, "Modelling the Correlation in the High-Frequency Noise of (Heterojunction) Bipolar Transistors using Charge-Partitioning", *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 221-224, 2003.
- [20] F. Bonani, S. Guerrieri, G. Ghione, "Compact conversion and cyclostationary noise modeling of pn-junction diodes in low-injection - Part I: model derivation", *IEEE Trans. Electron Devices*, Vol. 51, No.3, pp. 467-476, 2004.
- [21] J. Paasschens and R. de Kort, "Modelling the excess noise due to avalanche multiplication in (heterojunction) bipolar transistors", *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 108-111, 2004.
- [22] P. Sakalas, M. Ramonas, M. Schröter, C. Jungemann, A. Shimukovich and W. Kraus, "Impact Ionization Noise in SiGe HBTs: comparison of device and compact modeling with experimental results", *IEEE Trans. Electron Devices*, Vol. 56, No. 2, pp. 328-336, 2009.
- [23] G. Blasquez, J. Caminade, G. Le Gac, "Analysis of the effects of current crowding on noise of transistors with a circular geometry - application to transistor with any given geometry", *Physica B*, Vol. 92, pp. 313-329, 1977.
- [24] G. Blasquez, J. Caminade, G. Le Gac, "Experimental study of the effects of current crowding on noise of bipolar transistors at intermediate frequencies", *Physica B*, Vol. 94, pp. 359-365, 1978.
- [25] G. Blasquez, J. Caminade, K. van Vliet, "An accurate analysis of noise in rectangular bipolar transistors including current crowding", *Solid-State Electron*. Vol. 23, pp. 423-431, 1980.
- [26] J. C. J. Paasschens, "Compact Modeling of the Noise of a Bipolar Transistor Under DC and AC Current Crowding Conditions", *IEEE Trans. on Electron Dev.*, Vol. 51, No. 9, pp. 1483-1495, 2004.
- [27] M. Stoisiek and D. Wolf, "Origin of  $1/f$  noise in bipolar transistors", *IEEE Trans. on Electron Dev.*, Vol. 27, No. 9, pp. 1753-1757, 1980.
- [28] N. Siabi-Shahrivar, W. Redman-White, P. Ashburn, H. Kemhadijian, "Reduction of  $1/f$  noise in poly-silicon emitter bipolar transistors", *Solid-State Electron.*, Vol. 38, No. 2, pp. 389-400, 1995.
- [29] X.-Y. Chen, M. Deen, A. van Rheeën, Z.X. Yan, and M. Schröter, "Low-frequency noise in npn and pnp double-polysilicon BJTs, effects of temperature and emitter dimensions", *Proc. Int'l Electron Devices and Materials Symposium (IEDMS)*, 4 pages, 1998.

- [30] J. Deen, S. Rummyantsev, R. Bashir, R. Taylor, "Measurements and comparison of low frequency noise in npn and pnp polysilicon emitter bipolar transistors", *J. Applied Physics*, Vol. 84, No. 1, pp. 625-633, 1998.
- [31] J. Deen, S. Rummyantsev, M. Schröter, "On the origin of 1/f noise in polysilicon emitter bipolar transistors", *J. Applied Physics*, Vol. 85, No. 2, pp. 1192-1195, 1999.
- [32] J. Johansen, Z. Jin, J. Cressler, Y. Cui, G. Niu, Q. Liang, J.-S. Rieh, G. Freeman, D. Ahlgren, A. Joseph, "On the scaling limits of low-frequency noise in SiGe HBTs", *Solid-State Electron.*, Vol. 48, pp. 1897-1900, 2004.
- [33] E. Zhao, R. Krithivasan, A. Sutton, Z. Jin, J. Cressler, B. El-Kareh, S. Balster, H. Yasuda, "An Investigation of low-frequency noise in complementary SiGe HBTs", *IEEE Trans. Electron Devices*, Vol. 53, No.2, pp. 329-338, 2006.



## Chapter 8

### HICUM Level2

The theoretical background of compact model equations including the effects of bias, temperature, self-heating, noise, and lateral scalability, has been developed in the preceding chapters. This chapter describes the compact model HICUM/Level2 for bipolar transistors with mainly two objectives: (i) construction of the large-signal equivalent circuit (EC) with discussion on element formulation, and (ii) definition of model parameters and illustration of their impact on respective characteristics. The model parameters will be indicated in **bold** letters to make them better visible in the equations and the text. A complete list of model parameters is provided in section 8.7. The model equations are described on the basis of a vertical npn transistor. A vertical pnp transistor requires for most processes the addition of a parasitic n-well transistor through a subcircuit. A brief discussion of the small-signal model, which results from the derivatives of the large-signal model, is also provided along with the formulation of the included noise sources and their location in the EC. Many model equations will be visualized by displaying their general functional dependence, preferably in normalized form.

Figure 8.1 shows a schematic cross-section of a SiGe-HBT structure that serves as basis for the EC construction. The general requirements for a compact model, which were outlined in sec. 2.3, are being kept in mind while developing this model. The cross-section of Fig. 8.1 is partitioned into regions (indicated by polygons with different line styles) that are related to certain portions of the EC and, hence, lead to a physics-based EC. The portion in the dashed box under the emitter is called “internal transistor” that determines most of the fundamental characteristics. The other re-



lute values. Where applicable, corresponding corrections w.r.t. perimeter effects will be discussed.

Figure 8.2 shows the large-signal EC of the internal transistor. Branch currents resulting from nonlinear charges are indicated by time derivatives of the charges ( $\dot{Q}$ ). The EC contains the diode current  $i_{jBEi}$  and depletion charge  $Q_{jEi}$  of the BE junction, the diode current  $i_{jBCi}$  and depletion charge  $Q_{jCi}$  of the BC junction, and the voltage controlled transfer current source  $i_T$ . Furthermore, the minority charge is represented by its forward and reverse component,  $Q_f$  and  $Q_r$ . Moreover, breakdown in the BC SCR is taken into account by the avalanche current source,  $i_{AVL}$ , and possible tunneling through the BE SCR is represented by  $i_{BEti}$ . Finally, in SiGe HBTs the conduction band barrier forming at high current densities in the BC SCR leads to an additional recombination component  $i_{Bhrec}$ . The relation of these EC elements to distinctive regions in the transistor structure along with a physics-based description of the elements forms the core of a physics-based compact model.

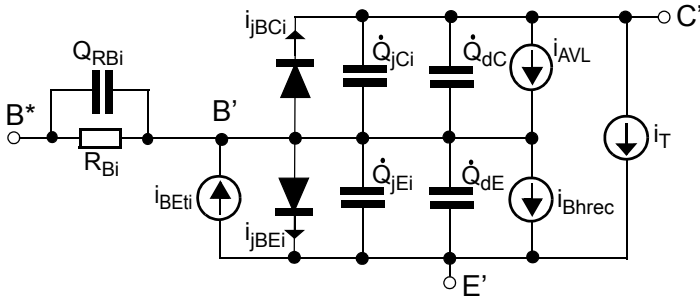


Fig. 8.2: Large-signal equivalent circuit of the internal transistor structure defined by the effective emitter area  $A_E$ .

Since for design and optimization of high-speed integrated bipolar circuits, accurate modeling of the dynamic transistor behavior is mandatory, in HICUM/L2 particularly those quantities, that determine the dynamic transistor behavior, such as depletion and minority charges, are regarded as basic variables of the model. From these charges the DC transfer and output characteristics can be described in a reasonably simple but accurate way by means of the generalized integral charge control relation (GICCR). Below, the model equations of HICUM/L2 are presented in a modular

way, starting with internal charges before the various currents are discussed. In the text below, the model parameters are defined at a reference device temperature  $T_0$ . For a simulation temperature different from  $T_0$ , their values need to be updated according to the temperature dependent equations outlined in section 8.3.

### 8.1.1 Depletion charges and capacitances

#### 8.1.1.1 Internal base-emitter junction

The model equation for the voltage ( $V_{B'E'}$ ) controlled internal BE depletion charge reads

$$Q_{jEi} = \frac{C_{jEi0} V_{DEi}}{1 - z_{Ei}} \left[ 1 - \left( 1 - \frac{v_j}{V_{DEi}} \right)^{(1 - z_{Ei})} \right] + a_{jEi} C_{jEi0} (V_{B'E'} - v_j). \quad (8.1)$$

Its derivative w.r.t.  $V_{B'E'}$  yields the internal BE depletion capacitance

$$C_{jEi} = \frac{C_{jEi0}}{(1 - v_j/V_{DEi})^{z_{Ei}}} \cdot \frac{dv_j}{dV_{B'E'}} + a_{jEi} C_{jEi0} \left( 1 - \frac{dv_j}{dV_{B'E'}} \right). \quad (8.2)$$

The parameter  $a_{jEi}$  is defined as the ratio of the maximum value to the zero-bias value. Above formulation is modified compared to the simple smoothing in (3.158) to avoid divergence problem from the negative derivative of the voltage controlled capacitance and to ease parameter extraction at higher forward bias. Keeping  $C_{jEi}$  constant at *very* high forward bias is also justified since in this operating region the current dependent diffusion capacitance is orders of magnitude larger than  $C_{jEi}$ . The auxiliary voltage  $v_j$  and its derivative were already given in sec. 3.4.2 but are repeated here for this specific case:

$$v_j = V_f - V_T \frac{x + \sqrt{x^2 + a_{ff}}}{2} < V_f, \quad (8.3)$$

with  $a_{ff} = 4 \ln^2(2) = 1.921812$  as constant<sup>1</sup>. The derivative in (8.2) is then given by

$$\frac{dv_j}{dV_{B'E}} = \frac{x + \sqrt{x^2 + a_{fj}}}{2\sqrt{x^2 + a_{fj}}} \text{ with } x = \frac{V_f - v_{B'E}}{V_T}. \quad (8.4)$$

The voltage

$$V_f = V_{DEi} [1 - a_{jEi}^{-(1/z_{Ei})}] \quad (8.5)$$

marks the intercept of the classical expression with the maximum value  $a_{jEi} C_{jEi0}$ .

Figure 8.3 shows the typical behavior of  $C_{jEi}$  as function of  $V_{B'E}$  for three sets of model parameters. Note that higher values of  $z_{Ei}$  increase the curvature, whereas higher values of  $V_{DEi}$  shift the curves towards higher forward voltages before they reach their maximum. The area under the CV curves in Fig. 8.3 corresponds to the charge given in (8.1).

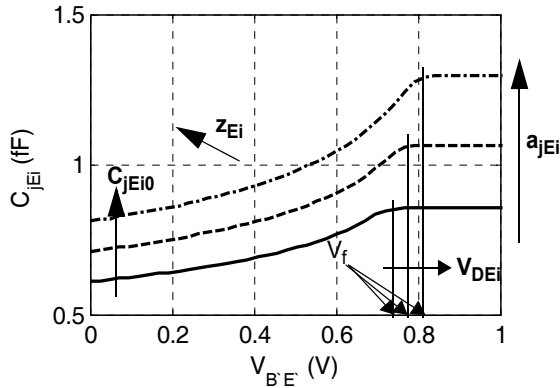


Fig. 8.3: Typical dependence of internal BE depletion capacitance on junction voltage in HICUM/L2. Arrows indicate the influence on characteristics for three sets  $\{T_1(-), T_2(--), T_3(-\cdot-)\}$  of parameters  $\{C_{jEi0}/\text{fF}, V_{DEi}/\text{V}, z_{Ei}, a_{jEi}\} = \{T_1(0.61, 0.88, 0.33, 1.4), T_2(0.71, 0.9, 0.42, 1.5), T_3(0.81, 0.92, 0.5, 1.6)\}$ .

1. The value results from a downward compatibility requirement after replacing the originally existing exponential smoothing function.

### 8.1.1.2 Internal base-collector junction

In contrast to the capacitance equation (3.163), HICUM/L2 uses a sophisticated formulation for all BC capacitance components in order to more accurately describe both high reverse bias and high-forward bias conditions. The internal BC depletion capacitance is modeled by

$$C_{jCi} = C_{jCi,cl} + C_{jCi,PT} + C_{jCi,fb}, \quad (8.6)$$

where the various components represent the different bias regions.

At at high forward bias the capacitance

$$C_{jCi,fb} = a_{jCi} C_{jCi0} \frac{1}{1 + e_{j,r}}, \quad (8.7)$$

is limited to its maximum value  $a_{jCi} C_{jCi0}$  analogous to  $C_{jEi}$ . The factor  $a_{jCi}$  ( $= 2.4$ ) is treated as constant rather than a model parameter since  $C_{jCi}$  is of little relevance at high forward or inverse operation.

At medium bias the classical equation is used,

$$C_{jCi,cl} = \frac{C_{jCi0}}{(1 - v_{j,m}/V_{DCi})^{z_{Ci}}} \cdot \frac{e_{j,r}}{1 + e_{j,r}} \frac{e_{j,m}}{1 + e_{j,m}}, \quad (8.8)$$

in which the two expressions at the end represent smoothing functions that provide a continuously differentiable transition to the adjacent regions of high forward and high reverse bias, respectively. Figure 8.4 shows a numerical example for the voltage dependence of the approach in (8.8).

Like for the BE depletion capacitance, the numerical overflow at high forward bias is avoided by replacing  $v_{B'C'}$  in the classical equation with the auxiliary (smoothed) voltage

$$v_{j,m} = -V_{jPCi} + V_r \ln[1 + e_{j,m}] \quad (8.9)$$

containing the terminal related punch-through voltage (cf. (3.162))

$$V_{jPCi} = V_{PTCi} - V_{DCi}, \quad (8.10)$$

with  $V_{PTCi}$  given by the physics-based expression (3.9). The voltage, at which the transition from medium to large reverse bias starts, is

$$V_r = 0.1 V_{jPCi} + 4 V_T. \quad (8.11)$$

These variables also enter the function

$$e_{j,m} = \exp\left(\frac{V_{jPCi} + v_{j,r}}{V_r}\right). \quad (8.12)$$

The additional auxiliary (smoothed) voltage

$$v_{j,r} = V_{fCi} - V_T \ln[1 + e_{j,r}] \quad (8.13)$$

avoids the pole at  $V_{DCi}$  and is controlled by the actual internal BC voltage via the function

$$e_{j,r} = \exp\left(\frac{V_{fCi} - v_{B'C'}}{V_T}\right). \quad (8.14)$$

Note, that  $v_{j,r}$  equals  $v_{B'C'}$  at large reverse bias. Following (8.5), the voltage defining the intercept between the classical expression and the maximum value at large forward bias for the internal BC junction is given by

$$V_{fCi} = V_{DCi} [1 - a_{jCi}^{-(1/z_{Ci})}]. \quad (8.15)$$

According to the above, “large reverse” bias is defined as  $v_{B'C'} \leq -V_{jPCi}$ , “medium” bias as  $V_{jPCi} < v_{B'C'} < V_{fCi}$ , and “large forward” bias as  $v_{B'C'} \geq V_{fCi}$ .

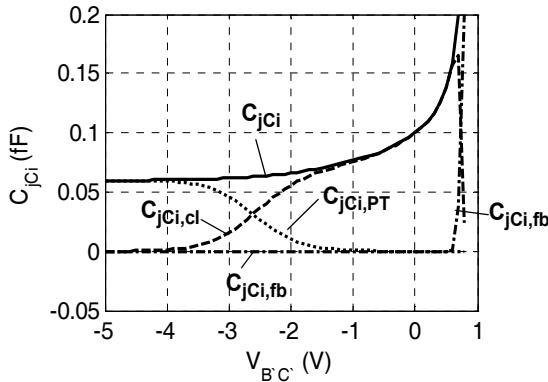


Fig. 8.4: Visualization of the different  $C_{jCi}$  components with the transition to the different bias regions:  $C_{jCi,cl}$ (--),  $C_{jCi,PT}$ (...),  $C_{jCi,fb}$ (-.), and total  $C_{jCi}$ (-).

The possible punch-through effect at large reverse bias is modeled by

$$C_{jCi,PT} = \frac{C_{jCi0,r}}{(1 - v_{j,r}/V_{DCi})^{z_{Ci,r}}} \cdot \frac{1}{1 + e_{j,m}}. \quad (8.16)$$

It employs the classical voltage dependence but with different parameters  $C_{jCi0,r}$  and  $z_{Ci,r}$ , which model the weak voltage dependence under punch-through condition (cf. Fig. 8.4) resulting from the “slower” penetration of the BC SCR into the highly doped base and buried layer region.  $C_{jCi0,r}$  can be calculated from the punch-through voltage as

$$C_{jCi0,r} = C_{jCi0} \cdot \left( \frac{V_{DCi}}{V_{PTCi}} \right)^{(z_{Ci} - z_{Ci,r})}. \quad (8.17)$$

Here,  $z_{Ci,r}$  is internally set to  $z_{Ci}/4$ , which turned out to be a reasonable choice after investigating many different doping profiles. Hence,  $C_{jCi0,r}$  and  $z_{Ci,r}$  are not needed as user adjustable model parameters.

Figure 8.5 shows  $C_{jCi}(V_{B'C'})$  from reverse to forward bias for three sets of model parameters to indicate their impact. Note that the effect of varying  $(C_{jCi0}, z_{Ci}, V_{DCi})$  is similar to varying  $(C_{jEi0}, z_{Ei}, V_{DEi})$  in the BE depletion capacitance. According to Fig. 8.5 the CV curve follows the classical behavior down to  $(V_{DCi} - V_{PTCi})$  at reverse bias and up to  $V_{fCi}$  at forward bias.

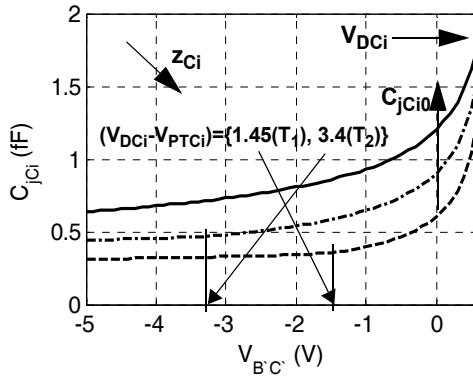


Fig. 8.5: Typical dependence of the internal BC depletion capacitance on internal junction voltage for three sets  $\{T_1(-), T_2(-), T_3(-)\}$  of parameters  $\{C_{jCi0}/\text{fF}, V_{DCi}/\text{V}, z_{Ci}, V_{PTCi}/\text{V}\} = \{T_1(0.6, 0.75, 0.5, 2.2), T_2(0.9, 0.8, 0.41, 4.2), T_3(1.2, 0.85, 0.33, 100)\}$  is indicated.



The charge expression is then obtained by integrating the internal BC depletion capacitance and reads

$$Q_{jCi} = Q_{jCi,m} + Q_{jCi,r} - Q_{jCi,c} + Q_{jCi,f} \quad (8.18)$$

with the respective components at medium bias,

$$Q_{jCi,m} = \frac{C_{jCi0} V_{DCi}}{1 - z_{Ci}} \left[ 1 - \left( 1 - \frac{v_{j,m}}{V_{DCi}} \right)^{1 - z_{Ci}} \right], \quad (8.19)$$

in the punch-through region,

$$Q_{jCi,r} = \frac{C_{jCi0,r} V_{DCi}}{1 - z_{Ci,r}} \left[ 1 - \left( 1 - \frac{v_{j,r}}{V_{DCi}} \right)^{1 - z_{Ci,r}} \right], \quad (8.20)$$

and at high forward bias,

$$Q_{jCi,f} = a_{jCi} C_{jCi0} (v_{B'C'} - v_{j,r}), \quad (8.21)$$

as well as with the “correction” component

$$Q_{jCi,c} = \frac{C_{jCi0,r} V_{DCi}}{1 - z_{Ci,r}} \left[ 1 - \left( 1 - \frac{v_{j,m}}{V_{DCi}} \right)^{1 - z_{Ci,r}} \right]. \quad (8.22)$$

At high current densities  $C_{jCi}$  becomes also current dependent, which was shown and investigated in sec. 4.4.1. As discussed in [2], for small-signal applications such as LNAs, a pure voltage dependent model for  $C_{jCi}$  proved to be sufficient, since transistors in (small-signal) analog circuits are not operated at high current densities. For large-signal transient applications the current dependence of  $C_{jCi}$  may become relevant for distortion, e.g., in power amplifiers. So far, a clear indication for the importance of this effect has been missing and, therefore, the current dependence of  $C_{jCi}$  is neglected in the present model version, partially also for numerical reasons.

### 8.1.2 Minority charge and transit times

The minority charge is divided into a “forward” component  $Q_f$  and a “reverse” (or inverse) component  $Q_r$ . These large-signal charge components are determined by integrating the respective small-signal transit times.

### 8.1.2.1 Forward minority charge component

The total operating point dependent mobile charge  $Q_f$  is fully assigned to the BE branch (cf. Fig. 8.2), i.e.

$$Q_{dE} = Q_f = Q_{f0} + \Delta Q_{fh}, \quad (8.23)$$

with  $Q_{f0}$  as low-current and  $\Delta Q_{fh}$  as high-current component, which will be discussed below in detail.

The low-current forward minority charge

$$Q_{f0} = \tau_{f0} i_{Tf} \quad (8.24)$$

depends linearly on the forward component  $i_{Tf}$  of the transfer current and the current independent low-current transit time  $\tau_{f0}$ .

Neglecting the second term in (3.237), and using (3.182), (3.204) and (3.222) for the remaining terms, the low-current component  $\tau_{f0}$  is modeled as function of the voltage  $V_{B'C'}$  [3]:

$$\tau_{f0}(V_{B'C'}) = \tau_0 + \Delta\tau_{0h}(c - 1) + \tau_{Bvl} \left( \frac{1}{c} - 1 \right). \quad (8.25)$$

Here,  $1/c = C_{jCi,t}(V_{B'C'})/C_{jCi0}$  is a normalized capacitance that is evaluated for the same model parameters as  $C_{jCi}$ , but with the punch-through voltage set to infinity. The corresponding stronger voltage dependence roughly models the impact of the bias dependent BC SCR moving stronger into the base under punch-through condition.

The first time constant in (8.25),  $\tau_0$ , represents the sum of the bias independent components of the various transistor regions at  $V_{B'C'} = 0$ . The second term in (8.25), introduced in (3.182), represents the overall voltage dependent change caused by both the Early-effect and the transit time through the BC space charge region (cf. Fig. 8.6). For  $\Delta\tau_{0h} < 0$  the reduction of the neutral base width through the Early-effect dominates while for  $\Delta\tau_{0h} > 0$  the transit time increase caused by the widening of the BC SCR at large voltages dominates. The third term corresponds to a simplified version of the first term in (3.182) and takes into account the finite carrier velocity in the BC SCR, resulting in a carrier jam at low  $V_{C'E'}$  voltages.

Figure 8.6 shows three examples for the voltage dependence of the low-current transit time. The case of  $T_3$ , with negative value of  $\Delta\tau_{0h}$ , represents the dominance of the Early effect. The case of  $T_1$ , with positive value of

$\Delta\tau_{0h}$ , describes the impact of the counteracting trends of the Early-effect at very low  $V_{C'E'}$  and of the BC SCR widening at larger reverse voltage  $V_{C'B'}$ . This leads to a minimum at  $V_{B'C'} = 0$  and an increase of  $\tau_{f0}$  towards reverse bias. Finally, the case of  $T_2$ , with  $\Delta\tau_{0h} = \tau_{Bvl} = 0$ , leads to a voltage independent  $\tau_{f0} = \tau_0$ .

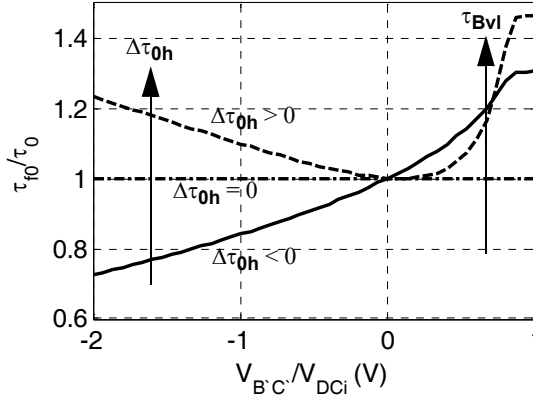


Fig. 8.6: Normalized low-current transit time as a function of normalized (internal) BC voltage shown for three examples  $T_1$ (--),  $T_2$ (-), and  $T_3$ (-) with parameter sets  $\{\Delta\tau_{0h}/\tau_0, \tau_{Bvl}/\tau_0\} = \{T_1(0.67, 0.61), T_2(0, 0), T_3(-0.33, 0.08)\}$ .

As discussed in sec. 5.3.4.2, the low-current transit time depends on the emitter dimensions. This is taken into account by applying (5.115) to the parameter  $\tau_0$  and letting  $\tau_{m0,A} \rightarrow \tau_{0i}$  and  $\tau_{m0,P} \rightarrow \tau_{0p}$  yielding

$$\tau_0 = \tau_{0i} \frac{1 + (\tau_{0p}/\tau_{0i})\gamma_C P_{E0}/A_{E0}}{1 + \gamma_C P_{E0}/A_{E0}} = \frac{\tau_{0i}}{f_{Qi}} \quad (8.26)$$

with  $\tau_{0i}$  and  $\tau_{0p}$ , respectively, as the bottom component and perimeter component, respectively, that are determined from measurements. Since the geometry calculation can be done by preprocessing, the actual HICUM/L2 model parameter reduces to  $f_{Qi}$ .

According to sec. 3.5.2.2 the onset of high current effects can be accurately characterized by the critical current  $I_{CK}$ . Combining (3.249) and (3.250) yields

$$I_{CK} = \frac{V_{ceff}}{r_{Ci0}} \frac{1}{[1 + (V_{ceff}/V_{lim})^{b_{ick}}]^{1/b_{ick}}} \left[ 1 + \frac{v + \sqrt{v^2 + a_{ickpt}}}{2} \right] \quad (8.27)$$

with  $v = (v_{ceff} V_{lim})/V_{PT}$  as argument, and  $a_{ickpt} = 10^{-3}$  as fixed value of the hyperbolic smoothing function that connects the low- and high-voltage regions. The internal collector voltage  $v_{ci} = v_{C'E'} - V_{CEs}$  given in (3.1) also needs to be smoothed properly by introducing the effective collector voltage

$$v_{ceff} = V_T \left[ 1 + \frac{u + \sqrt{u^2 + a_{vceff}}}{2} \right] \quad (8.28)$$

with  $u = (V_{ci} - V_T)/V_T$  and  $a_{vceff} = 1.921812$  as constant<sup>2</sup>. The model parameters  $r_{Ci0}$ ,  $V_{lim}$ ,  $V_{PT}$  and  $V_{CEs}$  have already been defined in sec. 3.5 and 3.1. Note that the low-field internal collector resistance  $r_{Ci0}$  does not include the impact of collector current spreading; the corresponding modifications for this effect are discussed in section 8.2.1. The parameter  $b_{ick}$  depends on the transistor type and usually equals 2 (npn) or 1 (pnp).

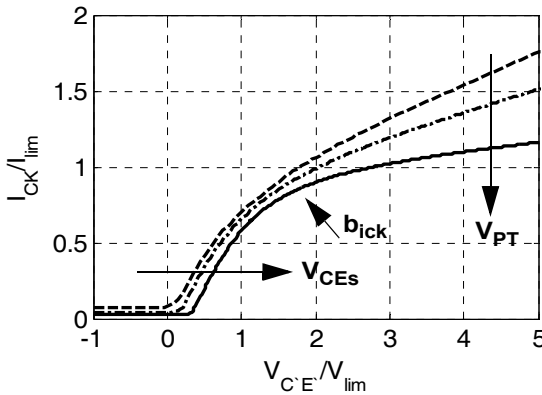


Fig. 8.7: Plots of normalized  $I_{CK}$  vs. normalized internal voltage  $V_{C'E'}$  for three different parameter sets:  $T_1$ (--),  $T_2$ (-), and  $T_3$ (.) with  $\{r_{Ci0}/\Omega, V_{lim}/V, V_{CEs}/V, V_{PT}/V\} = \{T_1(200, 0.4, 0.02, 2), T_2(150, 0.7, 0.1, 5), T_3(100, 1.0, 0.3, 20)\}$ , and  $b_{ick} = 2$ .

---

2. The value results from a downward compatibility requirement after replacing the originally existing exponential smoothing function.

Figure 8.7 shows the critical current for three different model parameter sets. Due to the normalization to  $I_{lim} = V_{lim}/r_{Ci0}$  the impact of the parameter  $r_{Ci0}$  is not visible. The normalization of  $V_{CE}$  to  $V_{lim}$  makes the impact of  $V_{lim}$  hardly visible, except for a small influence at high voltages, where the main variation is caused by  $V_{PT}$  though. Finally, the different values of  $V_{CEs}$  lead to the observed shift in the curves.

The nonlinear increase of the stored minority charge  $\Delta Q_{fh}$  toward high collector current densities is modelled in (8.23) by the regional components  $\Delta Q_{Ef}$ ,  $\Delta Q_{Bf}$  and  $\Delta Q_{Cf}$  as

$$\Delta Q_{fh} = \Delta Q_{Ef} + \Delta Q_{Bf} + \Delta Q_{Cf}. \quad (8.29)$$

According to (3.259), the additional charge stored in the neutral emitter is

$$\Delta Q_{Ef} = \Delta \tau_{Ef}(i_{Tf}) \frac{i_{Tf}}{1 + g_{\tau E}} \quad (8.30)$$

The bias-dependent emitter storage time is given by (3.258),

$$\Delta \tau_{Ef} = \tau_{Ef0} \left( \frac{i_{Tf}}{I_{CK}} \right)^{g_{\tau E}}, \quad (8.31)$$

with  $\tau_{Ef0}$  from (3.257) and  $g_{\tau E}$  as model parameters.

High-current effects in the collector cause both  $\Delta Q_{Bf}$  and  $\Delta Q_{Cf}$  to increase rapidly with current due to their common dependence on the electric field  $E_{jc}$  (cf. sec. 4.2). While in npn HBTs the barrier in the BC valence band suppresses  $\Delta Q_{Cf}$  until the conduction band barrier has formed, this charge is significant in BJTs and triggers the increase of  $\Delta Q_{Bf}$  through the component  $\Delta Q_{Bf,c}$ . In existing public domain HICUM versions, the physics-based but more complicated formulation using a collector field model (cf. sec. 4.2) has been replaced by a more simple and compact form, in which the normalized injection width

$$w = \frac{w_i}{w_{Ci}} = \frac{i + \sqrt{i^2 + a_{hc}}}{1 + \sqrt{1 + a_{hc}}} \quad (8.32)$$

is described directly as function of bias according to (4.7) and (4.141),

$$i = 1 - \left( \frac{I_{CK}}{i_{Tf}} \right)^{\beta_{Cbar}}. \quad (8.33)$$

The factors  $\alpha_{hc}$  and  $\beta_{Cbar}$  are model parameters. Since in BJTs the bias dependence of  $\Delta Q_{Bfc}$  and  $\Delta Q_{Cf}$  is very similar it can be described by (3.272), which reads in compact form

$$\Delta Q_{Cf} = Q_{pC} = \tau_{pCs} i_{Tf} w^2 \quad (8.34)$$

with the saturation storage time  $\tau_{pCs}$  from (3.274). By using in (8.34) the (hyperbolic) smoothing function (8.32) for  $w$  rather than for  $i_{Tf}$ , the collector charge becomes continuously differentiable over the whole bias region.  $\Delta Q_{Bf}$  is then described with the same bias dependent formulation,

$$\Delta Q_{Bf} = \Delta Q_{Bf,c} = \tau_{Bfvs} i_{Tf} w^2, \quad (8.35)$$

except that the different saturation storage time  $\tau_{Bfvs}$  from (3.278) is used.

As a consequence, for BJTs the two components were combined in earlier HICUM versions leading to the compact expression

$$\Delta Q_{fh,c} = \Delta Q_{Bf,c} + \Delta Q_{Cf} = \tau_{hcs} i_{Tf} w^2 \quad (8.36)$$

with the saturation time constant of base *and* collector,

$$\tau_{hcs} = \tau_{Bfvs} + \tau_{pCs}, \quad (8.37)$$

as model parameter that can be more easily determined by parameter extraction than the separate regional time constants. This formulation can still be used in existing versions. Furthermore, the ratio

$$f_{thc} = \tau_{pCs} / \tau_{hcs} \quad (8.38)$$

was introduced as a model parameter in order to allow bias dependent modeling of *collector* current spreading for geometry scaling. This way, the time constants in (8.34) and (8.35) can be calculated from the actual model parameters as

$$\tau_{pCs} = f_{thc} \tau_{hcs} \quad \text{and} \quad \tau_{Bfvs} = (1 - f_{thc}) \tau_{hcs}. \quad (8.39)$$

Above capability of separating the time constants turned out to be also useful for the HBT related extensions described below and, in addition, allows to partition the charges for the GICCR and the barrier related base current component.

In HBTs  $\Delta Q_{Cf}$  is significantly reduced as long as the potential barrier  $V_{Cbar}$  remains in the valence band but increases towards its homojunction value  $Q_{pC}$  from (8.34), if the bias dependent conduction band barrier (4.8),

$$\Delta V_{Cb}(i_{Tf}) = V_{Cbar} \frac{i + \sqrt{i^2 + a_{Cbar}}}{1 + \sqrt{1 + a_{Cbar}}}, \quad (8.40)$$

approaches  $V_{Cbar}$ . A rough approximation of this behavior is

$$\Delta Q_{Cf} \approx Q_{pC} \exp\left(\frac{\Delta V_{Cb}(i_{Tf}) - V_{Cbar}}{V_T}\right). \quad (8.41)$$

In (8.40) the barrier height  $V_{Cbar}$  is a model parameter, while  $a_{Cbar} = 0.01$  is a constant.

The minority charge stored in the base region of HBTs is according to (4.139)

$$Q_{Bf} = \left[ \tau_{Bfd} + \tau_{Bfv} \exp\left(\frac{\Delta V_{Cb}(i_{Tf})}{V_T}\right) \right] i_{Tf} \quad (8.42)$$

with the time constants defined along the same lines as in sec. 3.5 and [3],

$$\tau_{Bfd} = \frac{w_B^2(i_{Tf})}{F_\zeta \bar{\mu}_{nB} V_T}, \quad \tau_{Bfv} = \frac{w_B(i_{Tf})}{G_\zeta v_n(i_{Tf})} \quad (8.43)$$

and the barrier term (8.40).  $Q_{Bf}$  can be extended as follows,

$$Q_{Bf} = [\tau_{Bfd} + \tau_{Bfv}] i_{Tf} + \tau_{Bfv} i_{Tf} \left[ \exp\left(\frac{\Delta V_{Cb}(i_{Tf})}{V_T}\right) - 1 \right], \quad (8.44)$$

in order to separate the barrier related portion (second term) from the classical homojunction portion (first term). The latter also contains the low-current portion  $Q_{Bf0}$  that has already been included in  $Q_{f0}$ . The corresponding high-current portion of the first term in (8.44) equals  $\Delta Q_{Bf,c}$  only in BJTs, but in HBTs has to be split into an  $E_{jc}$  dependent component at medium current densities and a collector-triggered component along the lines of (8.35) at high current densities. This split is caused by the valence band barrier which reduces the injection width dependent portion significantly until the barrier builds up in the conduction band. For practical applications of HBTs the corresponding component of  $\Delta Q_{Bf,c}$  in the base,

like its collector counterpart (8.41), contributes only at very high current densities. Hence, they are lumped together for the same reasons as for (8.35):

$$\Delta Q_{fh,c} \approx \tau_{hcs} i_{Tf} w^2 \exp\left(\frac{\Delta V_{Cb}(i_{Tf}) - V_{Cbar}}{V_T}\right). \quad (8.45)$$

The  $E_{jc}$  dependent component of  $\Delta Q_{Bf,c}$  would have to be described separately by an  $E_{jc}$  model. Since this is too complicated for a compact model and the shape of the current dependent charge increase is mostly determined by the barrier related exponential term, the variable  $\tau_{BfV}$  is replaced by the constant  $\tau_{Bfvs}$  leading to the approximation:

$$\Delta Q_{Bf,b} \approx \tau_{Bfvs} i_{Tf} \left[ \exp\left(\frac{\Delta V_{Cb}(i_{Tf})}{V_T}\right) - 1 \right]. \quad (8.46)$$

The high-current charge formulation for HBTs then simply reads

$$\Delta Q_{Bf} + \Delta Q_{Cf} = \Delta Q_{fh,c} + \Delta Q_{Bf,b} \quad (8.47)$$

and allows to model barrier and high-current related storage effects separately. It reduces to the homojunction formulation for  $V_{Cbar} = 0$ .

The corresponding storage times are then given by the derivative of the charges with respect to  $I_{Tf}$  and can be calculated analytically. For instance, the increase of the transit time in BJTs is defined by (8.36) and reads

$$\Delta \tau_{fh,c} = \left. \frac{d\Delta Q_{fh,c}}{dI_{Tf}} \right|_{V_{CE}} = \tau_{hcs} w^2 \left[ 1 + \frac{2I_{CK}}{i_{Tf} \sqrt{i^2 + a_{hc}}} \right]. \quad (8.48)$$

For HBTs,  $\Delta Q_{fh,c}$  from (8.45) has to be inserted. Furthermore, with

$$\Delta \tau_{Bf,b} = \left. \frac{d\Delta Q_{Bf,b}}{dI_{Tf}} \right|_{V_{CE}} \quad (8.49)$$

the total transit time is then given by

$$\tau_f = \tau_{f0} + \Delta \tau_{fh} = \tau_{f0} + \Delta \tau_{Ef} + \Delta \tau_{Bf,c} + \Delta \tau_{Bf,b}. \quad (8.50)$$

Figure 8.8 shows the current dependence of the normalized *high-current* components  $\Delta \tau_{fh,c}$  and  $\Delta \tau_{Bf,b}$  for various model parameter sets. The parameters  $g_{fE}$ ,  $a_{hc}$ ,  $V_{Cbar}$  and  $\beta_{Cbar}$  are responsible for the curve shape,



i.e. the slope below and around  $I_{CK}$ . Note that the equations so far do not yet include the effect of collector current spreading.

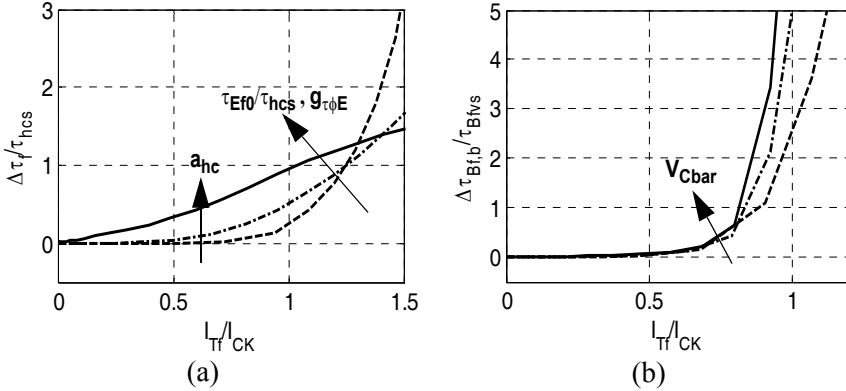


Fig. 8.8: Normalized transit time  $\Delta\tau_f$  vs. normalized forward transfer current  $I_{Tf}$  for three example sets of model parameters:  $T_1$ (--),  $T_2$ (-), and  $T_3$ (-) with  
 (a)  $\{\tau_{Ef0}/\tau_{hcs}, g_{\tau\phi E}, a_{hc}\} = \{T_1(0.1, 8, 0.05), T_2(0.2, 4, 0.3), T_3(0.5, 1, 0.7)\}$ ;  
 (b)  $\{V_{cbar}, \tau_{Bfvs}/ps\} = \{T_1(0.1, 5), T_2(0.15, 10), T_3(0.2, 50)\}$ .

### 8.1.2.2 Inverse minority charge component

At very low CE voltages or inverse operation, a current component  $i_{Tr}$  can be formally defined (cf. sec. 3.3.1 and 4.3.1) that is proportional to  $\exp(V_{B'C}/V_T)$ . The corresponding inverse minority charge  $Q_r$  is assigned to the BC diffusion charge  $Q_{dCi}$  in Fig. 8.2 and modeled as

$$Q_{dCi} = Q_r = \tau_r i_{Tr} \quad (8.51)$$

with  $\tau_r$  as inverse storage time.

### 8.1.3 Quasi-static transfer current

Following equation (4.87) and sec. 5.3.2 the total transfer current can be formally partitioned into a “forward” and “reverse” component

$$i_{Tf1} = \frac{c_{10}}{Q_{p,T}} \exp\left(\frac{v_{B'E}}{m_{cf} V_T}\right) \quad \text{and} \quad i_{Tr} = \frac{c_{10}}{Q_{p,T}} \exp\left(\frac{v_{B'C}}{V_T}\right). \quad (8.52)$$

Compared to the derivation of the GICCR in sec. 4.3.1, (8.53) and thus  $i_{TfI}$  have been extended by the non-ideality coefficient  $m_{Cf}$ . This factor takes into account transport effects such as thermionic emission and tunneling that may be significant mostly in III-V HBTs and that are not included in the GICCR derivation. Physically,  $i_{TfI}$  can be interpreted as the electron current flowing from emitter to collector at forward operation, while  $i_{Tr}$  can be interpreted as the electron current flowing from collector to emitter at inverse operation. This separation of  $i_T$  simplifies both the modelling of the minority charge components and the implementation of the nonlinear transfer current. The overall transfer current then reads

$$i_T = i_{TfI} - i_{Tr} = \frac{c_{10}}{Q_{p,T}} \left[ \exp\left(\frac{v_{B'E}}{m_{Cf} V_T}\right) - \exp\left(\frac{v_{B'C}}{V_T}\right) \right] \quad (8.53)$$

The total *weighted* sum of charges  $Q_{p,T}$  is given according to (4.84) by

$$Q_{p,T} = Q_{pj,T} + Q_{m,T}. \quad (8.54)$$

with the depletion charge related component

$$Q_{pj,T} = Q_{p0} + h_{jEi} Q_{jEi} + h_{jCi} Q_{jCi}. \quad (8.55)$$

The parameter  $Q_{p0}$  is the hole charge at thermal equilibrium;  $Q_{jEi}$  and  $Q_{jCi}$  are the depletion charge stored in the effective internal BE and BC junction, respectively;  $Q_{m,T}$  is the total weighted minority charge

$$Q_{m,T} = Q_{f0} + h_{fE} \Delta Q_{Ef} + \Delta Q_{Bf} + h_{fC} \Delta Q_{Cf} + Q_r, \quad (8.56)$$

with the low-current charge  $Q_{f0}$  from (8.24) and the high-current components  $\Delta Q_{Ef}$ ,  $\Delta Q_{Bf}$ ,  $\Delta Q_{Cf}$  and the inverse component  $Q_r$  from the previous section 8.1.2. According to sec. 4.3.1, the weight factors  $h_{jEi}$  and  $h_{jCi}$  as well as  $h_{fE}$  and  $h_{fC}$  represent the energy gap (or bandgap<sup>3</sup>) and mobility variations within the various transistor regions due to material composition and high-doping effects. For instance, in HBTs with compositional grading evaluation of the expressions from sec. 4.3.1 results in  $h_{jCi}$  being less than 1, which explains the larger Early voltages of those transistors as compared to transistors with a box-like composition. In public-domain HI-

---

3. Although the physically correct expression here is energy gap, the designation bandgap has been used in this chapter and other parts of the book since it has been in use for many years in the existing (engineering) literature.

CUM versions average values are assumed for the weight factors as a first-order approximation, but this may have to be extended towards a bias and temperature dependence for advanced SiGe HBTs with operating frequencies beyond 300 GHz. Note, that  $Q_{m,T}$  is generally *not equal to the actual stored minority charge*  $Q_m$ , that is used during dynamic operation.

In sec. 5.3.2 it was shown that the 1D GICCR can be extended to a 2D and 3D formulation to include the effects of emitter perimeter injection and collector current spreading. These effects generally (i.e. in all process generations known to the authors since the eighties) lead to a reduced current density in the internal transistor and, through the 2D weighting, to a decrease in  $Q_{p0h}$  at higher current densities as documented in Fig. 5.25. This in turn results in an increase of  $I_T$  compared to the 1D case. In early HICUM versions these effects have been lumped together in the simple expression

$$i_{Tf} = i_{Tf1}(1 + i_{Tf1}/I_{Ch}) \quad (8.57)$$

with  $i_{Tf1}$  from (8.52) and  $I_{Ch}$  as model parameter, which is (roughly) proportional to the emitter area. For infinite  $I_{Ch}$  the 1D formulation is obtained. Besides the lateral scalability of the model, the major advantages of this approach are that (i) a single equation can be used throughout the total operating region and (ii) a single transfer current source element can be used in the EC to describe 2D and 3D effects.

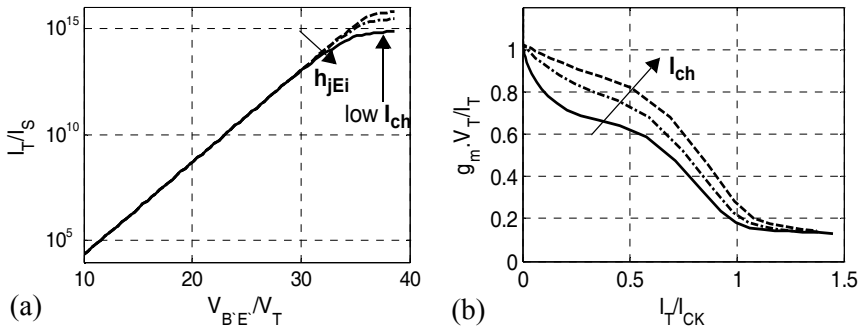


Fig. 8.9: (a) Normalized collector current ( $I_S = c_{I0}/Q_{p0}$ ) and (b) normalized transconductance for three sets of model parameters:  $T_1$ (--),  $T_2$ (-),  $T_3$ (-) with  $\{c_{I0}/fC^2s^{-1}, Q_{p0}/fC, I_{ch}/mA, h_{JEi}\} = \{T_1(5.4e-3, 6.8, 10, 10), T_2(5.4e-3, 3.1, 0.1, 1), T_3(5.4e-3, 0.8, 0.01, 0.1)\}$ . In all cases  $m_{Cf} = h_{JCI} = h_{fE} = h_{fC} = 1$ .

Figure 8.9 shows the normalized transfer current vs. the internal BE voltage at  $V_{B'C'} = 0V$  for three sets of model parameters, in which  $Q_{p0}$ ,  $h_{jEi}$ ,  $I_{Ch}$  have been varied. The impact of  $Q_{p0}$  and  $h_{jEi}$  is similar in changing the slope of the curve at low ( $Q_{p0}$ ) and medium ( $h_{jEi}$ ) current densities. A reduction of  $Q_{p0}$  also reduces the current at high current densities. Non-zero values of  $I_{Ch}$  increase  $I_T$  at high current densities.

At high reverse bias across either junction, the respective space-charge region can extend throughout the whole base region (base reach-through effect). As a result,  $Q_{p,T}$  in the GICCR would become zero or even negative, which would cause numerical problems. Physically, under such reach-through condition the current through the base is determined by thermionic emission across the base-emitter potential barrier. However, since this situation should not occur under any useful bias conditions and is therefore extremely unlikely with physics-based model parameters, the addition of a thermionic current component does not appear to be relevant from a computational and model application point of view. Such a current component and its derivatives would have to be numerically evaluated during all model calculations. Reach-through occurs most likely at very low or negative bias, where the (always positive) minority charge is negligible. Therefore, in HICUM the hole charge  $Q_{pj,T}$  is replaced by the smoothed charge

$$Q_{pT, low} = Q_{B, rt} \left( 1 + \frac{x + \sqrt{x^2 + a}}{2} \right), \quad (8.58)$$

with  $x = Q_{pj,T}/Q_{B,rt} - 1$  and  $a = 1.921812^4$ , which limits  $Q_{pj,T}$  to the positive value  $Q_{B,rt} = 0.05 Q_{p0}$ . For the usual operating range with  $Q_{pj,T}/Q_{p0} > 1$  the relation  $Q_{pj,T}/Q_{p,rt} \gg 1$  holds, so that  $Q_{p,low} - Q_{pj,T}$  is much smaller than  $10^{-6} Q_{p0}$ , and the computational effort associated with the smoothing can then be skipped.

In general, the GICCR is a nonlinear implicit equation for either  $i_T$  or  $Q_{p,T}$ . Since  $Q_{p,T}$  is the common variable in both current components  $i_{Tf}$  and  $i_{Tr}$ , the GICCR is solved for  $Q_{p,T}$  by employing Newton-Raphson iteration. However, at low current densities  $Q_{m,T}$  is a linearly varying func-

---

4. This value reproduces the results of the former exponential smoothing function.

tion of the current, i.e. the transit times are current *independent*, and  $I_{Ch}$  does not have any impact. Hence, the GICCR reduces to a quadratic equation with the explicit solution

$$Q_{p, T2} = \frac{Q_{p, Tlow}}{2} \left[ 1 + \sqrt{1 + 4c_{10} \frac{\tau_{j0} \exp\left(\frac{v_{B'E}}{V_T}\right) + \tau_r \exp\left(\frac{v_{B'C'}}{V_T}\right)}{Q_{p, Tlow}^2}} \right]$$

This solution may differ too much from the final solution at the transition from low to medium current densities since the transit times are continuous functions of currents. Therefore, a Newton iteration is always performed (except at inverse operation) using  $Q_{p, T2}$  for calculating a more general initial guess. For this,  $Q_{p, T2}$  is inserted into  $i_{Tf}$  and  $i_{Tr}$ . Then the resulting preliminary minority charge terms are added leading to

$$Q_{p, T, initial} = Q_{p, Tlow} + \tau_{j0} i_{Tf}(Q_{p, T2}) + \tau_r i_{Tr}(Q_{p, T2}). \quad (8.59)$$

This has turned out to be a very useful *initial guess* even at high current densities. At low and even medium current densities (below  $I_{CK}$ ) only 1 to 2 iterations are required to arrive at the final solution. Around  $I_{CK}$  the iteration count may increase to 5 depending on the current tolerance criterion of the simulator. It then decreases again to 2...3 at very high current densities.

### 8.1.4 Non-quasi-static effects

According to (4.204) the time dependent NQS transfer current can be described by a second-order differential equation,

$$A_2 \frac{d^2 i_{T, nqs}}{dt^2} + A_1 \frac{di_{T, nqs}}{dt} + i_{T, nqs} = i_T(t), \quad (8.60)$$

which relates the quasi-static transfer current  $i_T$  to its NQS counterpart  $i_{T, nqs}$  via the coefficients  $A_1$  and  $A_2$ . The latter result in principle from the original frequency-domain solution (4.203), in which  $A_1 = \tau_I$  and  $A_2 = \tau_1^2 - \tau_2^2$ . In [4] it was proposed to discretize (8.60) using the simulator time steps and the known QS transfer current. As the only difference the

coefficient  $A_2$  was changed to  $A_2 = \tau_1^2/3$  in order to obtain a Bessel polynomial, which is known from filter design to achieve a desired phase shift with minimum impact on magnitude. The latter requirement is not really necessary as was shown in sec. 4.5.4. But the small change of the coefficient  $A_2$  does not impact the accuracy and has the advantage that only the first-order time constant  $\tau_1$  is required as model parameter. Therefore and since the existing implementation code only needed to be modified slightly but did not require any changes to the equivalent circuit, this approach was employed in early HICUM versions (up to v2.1) and was applied to both the transfer current *and* the forward minority charge  $Q_f$ . For this, the corresponding first-order NQS time constants (a.k.a delay times),

$$\tau_{Qf} = \alpha_{Qf}\tau_f \quad \text{and} \quad \tau_{iT} = \alpha_{iT}\tau_f, \quad (8.61)$$

are modeled as function of the transit time  $\tau_f$  according to Fig. 4.49. The proportionality factors  $\alpha_{Qf}$  and  $\alpha_{iT}$  are model parameters. The approach in [4] has been working very well in practical circuit simulations.

In order to be able to use the general benefits of model compilers, the model equations have to be converted from FTN or C code to Verilog-A. Unfortunately, in Verilog-A access to the previous time steps is unavailable. Hence, starting with HICUM/L2 v2.21 the alternative was to realize the second-order differential equation through an adjunct LCR network or its corresponding gyrator equivalent [5], as shown in Fig. 8.10a for the example of the transfer current. The gyrator network avoids inductors, which is advantageous in certain circuit simulators. Unfortunately, a straight-forward use of the adjunct network in Fig. 8.10a is not possible since the bias dependence of the delay times<sup>5</sup> causes the model compiler to create undesired derivatives in the small-signal network. Although these derivatives are mathematically correct, they simply result from a physically inadequate representation of the NQS effect through the large-signal time domain adjunct network. For a bias dependent delay time, the additional derivatives were shown by device simulation to be non-physical.

An adjunct network implementation that avoids the undesired derivatives was proposed in [6]. If the equations of the gyrator network are modified as shown in Fig. 8.10b the corresponding results accurately match

---

5. In [5] a bias independent delay time is assumed.

both the small-signal and large-signal behavior obtained from Weil's approach. The same implementation, which has been available from version 2.23 on, is being used for the corresponding minority charge related adjunct network. It is important to mention that the original HICUM/L2 equivalent circuit has not been changed at all; i.e. the two adjunct networks have been added separately and are basically Verilog-A artifacts from the implementation point of view for NQS effects. They will only be evaluated if NQS effects are turned on using the parameter flag *flnqs*.

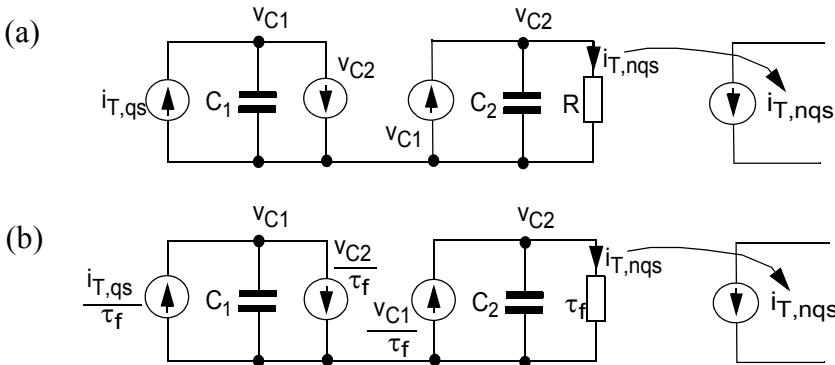


Fig. 8.10: (a) Adjunct gyrator equivalent of LCR circuit with  $C_1 = \alpha_{iT}\tau_f$ ,  $C_2 = C_1/3$ ,  $R = 1\Omega$ . (b) Gyrator equivalent circuit with normalization to the transit time in order to eliminate the undesired derivatives due to the bias dependent delay time. A similar subcircuit holds for the minority charge  $Q_f$ .

### 8.1.5 Quasi-static base current components

According to (3.101), (3.113), and (4.92a) the quasi-static current flowing from base to emitter across the bottom emitter area consists of back-injection and recombination components and is given by

$$I_{jBEi} = I_{BEiS} \left[ \exp\left(\frac{V_{B'E}}{m_{BEi}V_T}\right) - 1 \right] + I_{REiS} \left[ \exp\left(\frac{V_{B'E}}{m_{REi}V_T}\right) - 1 \right] \quad (8.62)$$

with  $I_{BEiS}$  and  $I_{REiS}$  as saturation currents. The ideality coefficient  $m_{BEi} \geq 1$  represents second-order effects such as Auger recombination, the (very small) modulation of the neutral emitter width, and the influence of the emitter poly-silicon or metal contact. The parameter  $m_{REi}$  is used to take into account the various volume recombination levels and effects in

the BE SCR. It usually assumes values in the range of 1.5 to 2.5. The equation is applicable up to high frequencies. Note though that at high switching speeds or frequencies the dynamic (i.e. capacitive) component of the terminal related base current becomes much larger than above quasi-static component.

The base current component flowing across the internal BC junction can be described in a similar way but is simplified to

$$I_{jBCi} = I_{BCiS} \left[ \exp\left(\frac{V_{B'C}}{m_{BCi}V_T}\right) - 1 \right] \quad (8.63)$$

with the saturation current  $I_{BCiS}$  and ideality coefficient  $m_{BCi}$  as model parameters. This component becomes significant only at hard-saturation or inverse operation.

In SiGe HBTs the valence band barrier leads to excess minority charge storage and associated recombination in the neutral base. This can be described by (4.93),

$$I_{Bhrec} = \Delta Q_{Bf} / \tau_{Bhrec}, \quad (8.64)$$

with  $\tau_{Bhrec}$  as geometry independent model parameter and  $\Delta Q_{Bf}$  as the additional minority charge in the base given by the sum of  $\Delta Q_{Bf,c}$  and  $\Delta Q_{Bf,b}$  in section 8.1.2.1.

Figure 8.11 shows an example of the various  $V_{B'E'}$ -controlled base current components at forward bias.

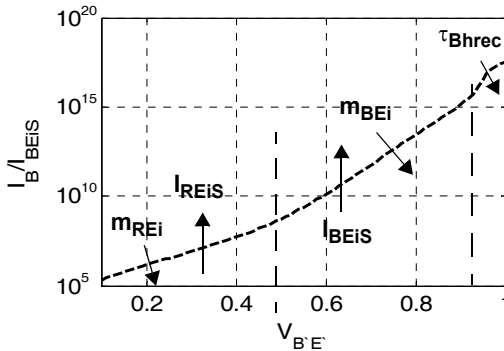


Fig. 8.11: Illustration of the various forward base current components vs. internal BE voltage with  $\{I_{BEiS}/A, m_{BEi}, I_{REiS}/A, m_{REi}, \tau_{Bhrec}/ps\} = (22.5e-21, 1, 10e-16, 2.2, 7)$ .



### 8.1.6 Collector-base breakdown

As discussed in sec. 4.3.3 the avalanche process results fundamentally from a non-local effect the impact of which though is small for wider collector regions ( $w_{Ci} > 0.3\mu\text{m}$ ) like those in high-voltage transistors. An accurate description including non-local effects requires the (bias dependent) modeling of the electric field within the BC SCR. Suitable formulations of a field model, such as those in [7] or in sec. 4.2 are under investigation but not yet available for reliable simulation. Therefore, the existing HICUM versions still contain the avalanche current model (3.133) that was derived from a local field approximation. To make the expression numerically stable, the term  $(1 - V_B' C' / V_{DCi})$  is replaced by the corresponding normalized depletion capacitance  $c = C_{jCi0} / C_{jCit} (v_B' C')$  that is also used in (8.25) for  $\tau_{ff}$ . Then, the constants in (3.133) can be lumped together into the avalanche effect related model parameters

$$f_{AVL} = \frac{2a_n}{b_n} \quad \text{and} \quad q_{AVL} = \frac{b_n \epsilon}{2} A_E, \quad (8.65)$$

which leads to the model formulation

$$I_{AVL} = I_{Tf} f_{AVL} V_{DCi} c^{1/z_{Ci}} \exp\left(-\frac{q_{AVL}}{C_{jCi0} V_{DCi} c^{1/z_{Ci}-1}}\right). \quad (8.66)$$

The terms  $f_{AVL} V_{DCi}$  and  $q_{AVL} / (C_{jCi0} V_{DCi})$  have not been merged into single parameters due to their different temperature dependence. This model is intended to indicate the onset of breakdown, i.e. of weak avalanche, under the emitter (planar breakdown). Thus,  $q_{AVL}$  scales directly with  $A_E$ . In favor of a numerically stable and continuously differentiable formulation, (8.66) does not include the influence of mobile charge on the electric field distribution in the collector at high current densities.

Figure 8.12 shows  $I_{AVL}$ , normalized to  $I_T$ , as function of the internal BC voltage for different values of  $f_{AVL}$  and  $q_{AVL}$ . The latter dominates the behavior since it occurs in the exponential function. Generally, the lower  $q_{AVL}$  the higher the avalanche current.

At high current densities, the electric field in the collector changes its shape and builds up again at the buried layer. This region is of little interest though for circuit design due to the very large charge storage and low

speed of the transistor in this bias region. Furthermore, usually the 3D pinch-in effect already occurs. Therefore, lumped modeling of the *current dependent* avalanche breakdown is not a high priority since (i) the bias region is not practically relevant, (ii) it increases the computational effort and reduces the numerical stability of a model, and (iii) capturing the 3D current pinch-in effect is not possible in any way.

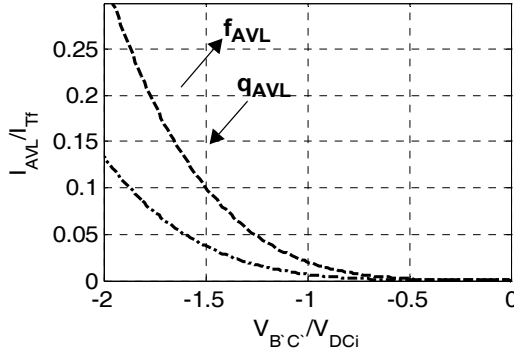


Fig. 8.12: Normalized avalanche current  $I_{AVL}/I_{TF}$  as a function of normalized internal BC voltage  $V_{BC}/V_{DCi}$  for two sets  $T_2$ (-- ) and  $T_3$ (- ) with parameters  $\{C_{jCi0}/\text{fF}, q_{AVL}/\text{fC}, f_{AVL}/V\} = \{T_2(0.94, 11, 900), T_3(0.94, 12, 800)\}$ .

### 8.1.7 Emitter-base tunneling

In most SiGe (and also III-V) processes tunneling occurs at the internal (bottom) BE junction. The present model version includes band-to-band tunneling based on (4.102) and (4.103). In order to make these equations suitable for a compact model,  $w_{BE} = \varepsilon/C_{jEi}$  is used in  $E_j$  from (4.103). Inserting  $E_j$  into (4.102), introducing the normalized voltage  $v_e = v_{B'E'}/V_{DEi}$ , and realizing that  $(1 - v_e) = c_e^{-1/z_{Ei}}$  yields

$$i_{btb} = I_{btbS}(-v_e)c_e^{1-1/z_{Ei}}\exp[-a_{btb}c_e^{1/z_{Ei}-1}] \quad (8.67)$$

with the model parameters  $I_{btbS}$  and  $a_{btb}$  that are given by

$$I_{btbS} = \sqrt{\frac{8qm^*q^2V_{DEi}^2}{V_g h^2 \varepsilon}} C_{jEi0}, \quad a_{btb} = \frac{A_E \varepsilon}{C_{jEi0}} \frac{4\pi \sqrt{2m^*qV_g V_g}}{3hV_{DEi}}. \quad (8.68)$$

Expressing  $v_e$  by  $c_e$  makes the formulation continuously differentiable via the smooth depletion capacitance expression. Figure 8.13 shows an exam-

ple for the bias dependence of the model equation with different parameter values. Band-to-band tunneling dominates at reverse bias but is negligible at forward bias where other mechanisms, such as trap-assisted tunneling can become important. In the present model the total tunneling current is given by  $I_{BEt} = I_{btb}$ .

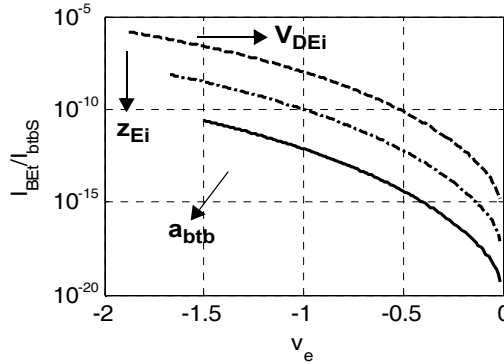


Fig. 8.13: Normalized tunneling current as a function of normalized internal BE voltage  $v_e$  for different sets  $T_1(-)$ ,  $T_2(-)$ , and  $T_3(-)$  of parameters  $\{a_{btb}, V_{DEi}/V, z_{Ei}\} = \{T_1(30, 0.8, 0.33), T_2(35, 0.9, 0.42), T_3(40, 1, 0.5)\}$ .

### 8.1.8 Internal base resistance

The bias and geometry dependent quasi-static internal base resistance formulation includes the effects of conductivity modulation and DC emitter current crowding. By combining the results (5.31) and (5.151), and following (5.166) one obtains

$$R_{Bi} = R_i \frac{\ln(1 + \eta)}{\eta} g_p. \quad (8.69)$$

Compared to (5.166) the perimeter charge storage related function  $g_p$  has been added that will be discussed later.  $R_i$  represents the conductivity modulation employing (5.31) by introducing

$$Q_{p0r} = (1 + f_{DQr0}) Q_{p0} \quad (8.70)$$

with  $f_{DQr0}$  as model parameter. However, (5.31) also requires the calculation of the internal charge  $\Delta Q_p = Q_{jEi} + Q_{jCi} + Q_f + Q_r$ . Depending on the model parameters,  $Q_{p0r} + \Delta Q_p$  can become zero or negative for a reverse

biased transistor or even just a BE junction<sup>6</sup>. Therefore, defining the normalized charge

$$q_{cm} = 1 + \frac{Q_{jEi} + Q_{jCi} + Q_f + Q_r}{Q_{p0r}} \quad (8.71)$$

and the smoothed expression

$$R_i = R_{Bi0} \frac{2}{q_{cm} + \sqrt{q_{cm}^2 + a_{cm}}} \quad (8.72)$$

with  $a_{cm} = 0.01$  avoids a possible numerical overflow.  $R_{Bi0}$  is a (geometry dependent) model parameter.

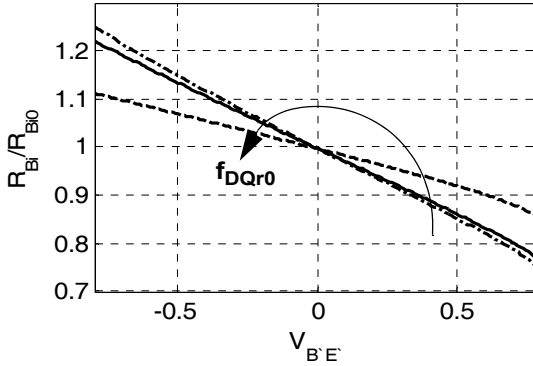


Fig. 8.14: Typical bias dependence of the normalized internal base resistance at low injection showing the effects of parameter variation for different sets:  $T_1$ (--),  $T_2$ (-.), and  $T_3$ (-) of parameters  $\{f_{DQr0}, f_{geo}\} = \{T_1(0, 0.67), T_2(0.1, 0.67), T_3(0.2, 0.67)\}$ .

Following (5.158a) the DC emitter current crowding factor is modeled by,

$$\eta = f_{geo} \frac{R_i i_{BEi}}{V_T}, \quad (8.73)$$

with  $f_{geo} = 1/(g_\eta g_i)$  as model parameter. Figure 8.14 shows the typical bias dependence of the (normalized) internal base resistance for different parameter sets.

---

6. This condition may also happen unintentionally during the Newton iteration of the circuit simulator.

In (8.69) the function

$$g_p = \frac{Q_{jEi} + f_{Qi} Q_f}{Q_{jEi} + Q_f} \quad (8.74)$$

is employed that is supposed to take into account the impact on  $R_{Bi}$  of the *transformation* of the minority charge storage from its original location at the emitter perimeter to the effective internal transistor [2]. This transformation actually can only be performed for the small-signal case, and (8.74) is a heuristic generalization for the large-signal case. At the time (8.74) was formulated the model equations were still coded by hand and allowed to retain the original expression

$$g_p = \frac{C_{jEi} + f_{Qi} C_f}{C_{jEi} + C_f} \quad (8.75)$$

for the small-signal case. With the use of Verilog-A code and model compilers only (8.74) is implemented while for the small-signal case its derivative is automatically generated. Since (i) this derivative can become both very different from the original expression (8.75) and computationally expensive, and (ii) in modern HBTs the area-specific depletion capacitance increases and the transit time decreases, it is recommended to drop  $g_p$  from (8.69) or set it to 1 by setting  $f_{Qi} = 1$ .

At high frequencies (or during fast transients) *dynamic emitter current crowding* causes a reduction of the impedance seen into the internal base node compared to DC or low-frequency conditions. A detailed discussion of this effect, a.k.a. *lateral non-quasi-static* effect, was given in sec. 5.2. Unfortunately, the exact solution (5.44) for the small-signal internal base impedance leads to a form that cannot be used in a compact model since its compact time-domain representation does not exist. However, further investigations in sec. 5.2.2 showed that the first-order series expansion (5.50) (or (5.165) for the 3D case) and subsequent conversion to (5.51) provides a reasonably accurate representation of the exact solution. The advantage of (5.51) is that it can also be directly converted to an equivalent circuit consisting of the DC internal base resistance  $R_{Bi}$  in parallel with the capacitance

$$C_{rBi} = f_{CrBi} (C_{jEi} + C_{jci} + C_{dEi} + C_{dCi}). \quad (8.76)$$

The model parameter  $f_{CrBi}$  depends on the emitter geometry according to (5.164). The internal diffusion capacitances are obtained as

$$C_{dEi} + C_{dCi} = f_{Qi}(C_{dE} + C_{dC}). \quad (8.77)$$

However, since this small-signal solution cannot be implemented in Verilog, it is recommended to set  $f_{Qi} = 1$ . The derivation of (8.76) is based on the assumption of negligible DC emitter current crowding. Therefore, if this assumption is violated, (8.76) must not be used. This also applies to fast *large-signal* transient applications, during which strong transient emitter current crowding in general occurs.

Since the theory leading to (8.76) is entirely based on small-signal considerations it does not provide any clue regarding the (large-signal) charge stored on the capacitor  $C_{rBi}$ . An analytical description of the internal base resistance for the high-speed large-signal transient mode is still an unresolved problem as was elaborately discussed in sec. 5.2.3. Since modern model implementation using Verilog-A and model compilers do not allow anymore to implement  $C_{rBi}$  directly in the *small-signal* code section, but require the associated (*large-signal*) charge to be implemented, the small-signal EC representation using just (8.76) cannot be realized anymore. The best approximation so far has been to define the corresponding nonlinear charge element as

$$Q_{rBi} = C_{rBi} V_{B^*B'}, \quad (8.78)$$

which produces the desired capacitance element in parallel to  $R_{Bi}$ . However, due to the bias dependence of  $C_{rBi}$  according to (8.76), additional undesired derivatives (i.e. transcapacitances) are automatically generated by a model compiler. It may be possible to circumvent this problem by a suitable adjunct network at the expense of an increase in computational effort. As a consequence, it is strongly recommended not to use  $C_{rBi}$  for any type of *transient* analysis and to consider  $C_{rBi}$  only as a means for improving the accuracy for *small-signal* circuit design.

## 8.2 Complete transistor

The model for the effective internal transistor described above is now extended by the elements describing the remaining regions of real transistor structures such as the one sketched in Fig. 8.1, consisting of the emitter perimeter region as well as the external and parasitic transistor regions. Furthermore, certain equations of the internal transistor need to be extended to include additional 2D and 3D effects. The electrical large-signal EC of the complete transistor is given in Fig. 8.15. Based on the results of chapter 5, the various elements will be discussed in the following sections. The absolute values of the EC elements need to be either extracted directly or calculated from a geometry scaling preprocessor. For instance, the program in [8] includes an extensive set of scaling formulas for different technologies and has been used extensively to generate model parameters for a large variety of transistor configurations.

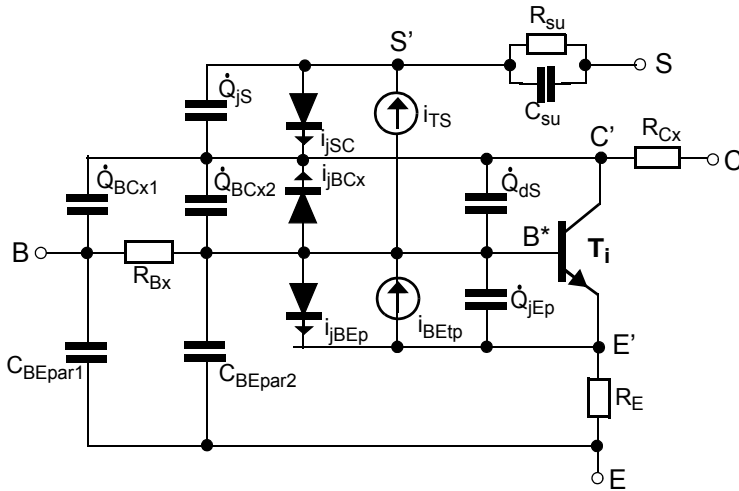


Fig. 8.15: Complete electrical large-signal EC of HICUM/L2 with the internal transistor  $T_i$  defined in Fig. 8.2.

Note that in Fig. 8.15 additional adjunct networks representing self-heating, the vertical NQS effect in  $I_T$  and  $Q_f$ , and correlated HF noise have been omitted yet. They are discussed separately and can also be realized directly in the code, i.e. without adjunct networks.

### 8.2.1 Mobile charge including collector current spreading

As shown in sec. 5.4.2, transistors with emitter dimensions that are in the order of the collector width  $w_{Ci}$  can, depending on the lateral extension of the SIC, exhibit significant current spreading in the collector. The result is an extension of mobile charge into the external collector region and a reduction of the transfer current density in the internal collector region (under the emitter). The latter shifts the onset of high-current effects to larger values of the transfer current and can be described by a modification of the critical current  $I_{CK}$ . The spreading of the mobile charge leads to a geometry dependent bias dependence of the high-current components of  $Q_f$ .

According to (5.128) current spreading increases the critical current of the internal transistor, defined without current spreading as  $J_{CK}A_E$ , by a collector current spreading factor (5.131),

$$f_{cs} = \begin{cases} \frac{lat_b - lat_l}{\ln[(1 + lat_b)/(1 + lat_l)]} & , lat_b > lat_l \\ 1 + lat_b & , lat_b = lat_l \end{cases} \quad (8.79)$$

It depends on collector width, emitter dimensions and current spreading angle  $\delta_C$  through the model parameters

$$lat_b = 2 \frac{w_{Ci}}{b_E} \tan \delta_C \quad \text{and} \quad lat_l = 2 \frac{w_{Ci}}{l_E} \tan \delta_C, \quad (8.80)$$

which correspond to  $\zeta_b$  and  $\zeta_l$ , respectively, in (5.132). In practice, their values are calculated by a geometry scaling preprocessor, since they do not depend on bias or temperature. The parameters will also be needed later on for the charge model. According to (5.128), the factor  $f_{cs}$  can be included in the already existing parameter  $r_{Ci0}$  as

$$r_{Ci0} = \frac{r_{Ci0,A}}{A_E} \frac{1}{f_{cs}}, \quad (8.81)$$

with  $r_{Ci0,A}$  as area specific 1D value. Above parameter is then inserted in (8.27) to calculate the bias dependent value of  $I_{CK}$  in the model.



Current spreading can be included in both the base and the collector minority charge through the modified  $I_{CK}$  above and also, according to (5.133), a modification of the normalized injection width

$$w = \begin{cases} \frac{\kappa - 1}{lat_l - \kappa lat_b} & , \quad lat_b > lat_l \\ \frac{1}{lat_b} \left[ \frac{1 + lat_b}{1 + i_\kappa lat_b} - 1 \right] & , \quad lat_b = lat_l \end{cases} \quad (8.82)$$

with

$$\kappa = \frac{1 + lat_l}{1 + lat_b} \exp \left[ i_\kappa \ln \left( \frac{1 + lat_b}{1 + lat_l} \right) \right] = \left( \frac{1 + lat_b}{1 + lat_l} \right)^{i_\kappa - 1} \quad (8.83)$$

Here, the normalized current is given by

$$i_\kappa = 1 - \frac{i + \sqrt{i^2 + a_{hc}}}{1 + \sqrt{1 + a_{hc}}} \quad (8.84)$$

with  $i$  from (8.33). Above formulations now replace (8.32).

The base charge component is then being calculated as

$$\Delta Q_{Bf,c} = \tau_{Bfvs} i_{Tf} w^2 \exp \left( \frac{\Delta V_{Cb}(i_{Tf}) - V_{Cbar}}{V_T} \right) \quad (8.85)$$

in which  $\tau_{Bfvs}$  is given by (8.39). The barrier related component  $\Delta Q_{Bfb}$  is still calculated from (8.46). The geometry dependent  $I_{CK}$  enters both charge expressions via the normalized injection width and the barrier height (8.40).

However, beyond  $I_{CK}$ , i.e. once an injection width forms, the collector component  $\Delta Q_{Cf}$  needs to be treated differently from the case of negligible current spreading. Following (5.136) and (5.137) it is given by

$$\Delta Q_{Cf} = \tau_{pCs} I_{Tf} \begin{cases} 2 \frac{f_{Ci} \ln \left( \frac{1 + lat_b^w}{1 + lat_l^w} \right) - f_{Cb} + f_{Cl}}{lat_b - lat_l} & , \quad lat_b > lat_l \\ \frac{1 + lat_b^{w/3}}{1 + lat_b^w} w^2 & , \quad lat_b = lat_l \end{cases} \quad (8.86)$$

with  $\tau_{pCs}$  from (8.39). Since the expression after the parenthesis depends on bias through  $w$ , a geometry dependent preprocessing is not possible in this case. The auxiliary bias-dependent functions

$$f_{Ci} = w + \frac{\mathbf{lat}_b + \mathbf{lat}_l}{2} w^2 + \frac{\mathbf{lat}_b \mathbf{lat}_l}{3} w^3 \quad (8.87)$$

and

$$f_{Cb} = \frac{1}{\mathbf{lat}_b} \left( 1 - \frac{\mathbf{lat}_l}{\mathbf{lat}_b} \right) \left[ \frac{x^2 [2 \ln x - 1] + 1}{4} \right] + \frac{1}{\mathbf{lat}_b \mathbf{lat}_b} \left[ \frac{x^3 [3 \ln x - 1] + 1}{9} \right] \quad (8.88)$$

with  $x = 1 + \lambda \alpha \tau_b w$ , as well as  $f_{Cl} = f_{Cb}(\mathbf{lat}_b \leftrightarrow \mathbf{lat}_l)$  depend not only on geometry but also on bias and temperature. In the actual implementation of these equations possible divisions by zero that may occur for  $\mathbf{lat}_b \rightarrow 0$  or  $\mathbf{lat}_l \rightarrow 0$  have been taken into account by appropriate series expansions. In case of negligible current spreading ( $\mathbf{lat}_b = \mathbf{lat}_l = 0$  or  $f_{thc} = 0$ ) above equations should not be evaluated at all.

The complete analytical formulation for 3D collector current spreading described above consists of a bias and temperature dependent portion, which needs to be evaluated during model runtime, and a just geometry dependent portion that is included through pre-processing in  $r_{Ci0}$  and hence  $I_{CK}$ . The bias dependent portion obviously adds a significant number of arithmetic operations. In order to speed up model execution, a simplification is proposed below, that maintains sufficient accuracy in the bias region of interest. The idea of the proposed solution is based on a shift of data processing from within the simulator to a geometry preprocessing tool with corresponding model parameter modifications.

During the preprocessing, the complete set of 3D collector current spreading equations described above is exercised only at  $I_{Tf} = I_{CK}$ , resulting in the transit time  $\tau_f(I_{CK})$ . The latter is supposed to closely match the reference data (i.e. measurements), since its model parameters have been extracted from that same data. On the other hand, including current spreading only in  $I_{CK}$  but setting  $\mathbf{lat}_b = \mathbf{lat}_l = 0$  in  $w$  of (8.82)-(8.84) and  $\Delta Q_{CF}$  of (8.86)-(8.88), i.e. exercising the 1D formulation given in section 8.1.2.1, yields the (generally larger) transit time  $\tau_{f,ncs}(I_{CK})$ . From these two time constants, a correction factor can be found,

$$f_{cst} = \frac{\tau_f(I_{CK}) - \tau_{f0}}{\tau_{f,ncs}(I_{CK}) - \tau_{f0}}. \quad (8.89)$$

This factor can be used to adjust the transit time of the 1D formulation such that its transit time matches the actual one at  $I_{CK}$ .

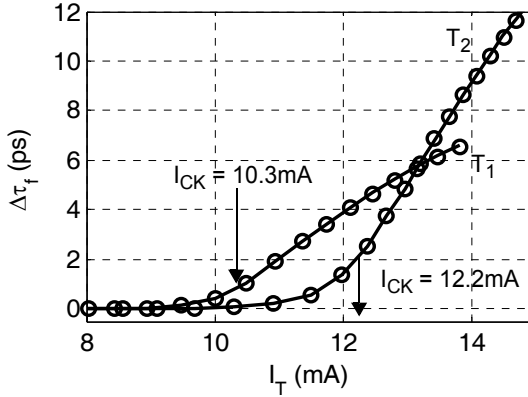


Fig. 8.16: Transit time vs. transfer current for different internal CE voltages for two sets of parameters  $\{\tau_{hcs}/ps, a_{hc}\} = \{T_1(20, 0.01), T_2(50, 0.05)\}$ : complete 3D collector current spreading equations with  $lat_b = 5.55$  and  $lat_l = 0.55$  (solid lines), 1D formulation with  $I_{CK}(f_{cs})$  and correction factor  $f_{cst}$  (symbols). The emitter transit time was turned off in this example to increase the sensitivity w.r.t. inaccuracies.

Figure 8.16 shows a comparison of the various curves for an example with a large current spreading angle and a large relative contribution of the collector transit time. The corrected curves match the reference curve very well in the region of the increase below and up to the critical current  $I_{CK}$ , and show still an acceptable accuracy well beyond  $I_{CK}$ . Thus, the computationally more expensive 3D current spreading calculations can often be replaced by exercising the simpler 1D formulation with just the 3D  $I_{CK}$  and a modified model parameter

$$\tau_{hcs} = f_{cst}\tau_{pCs} + \tau_{Bfvs}. \quad (8.90)$$

In practice, be  $\tau_{hcs,x}$  the parameter that is extracted from measurements together with the partitioning factor  $f_{thc}$  from (8.38) and the collector current spreading angle, then the actual model parameter inserted in the model card is calculated as

$$\tau_{hcs} = [(1 - f_{thc}) + f_{cst} f_{thc}] \tau_{hcs, x}, \quad (8.91)$$

where (8.39) was used.

### 8.2.2 Emitter perimeter region

The emitter perimeter includes both the junction region in the silicon (as discussed in sec. 5.3) and the BE spacer related isolation above the silicon. The associated components in Fig. 8.15 are discussed below.

According to (5.105) and (5.106) the perimeter base current can be described as

$$I_{jBEp} = I_{BEpS} \left[ \exp \left( \frac{V_{B^*E'}}{m_{BEp} V_T} \right) - 1 \right] + I_{REpS} \left[ \exp \left( \frac{V_{B^*E'}}{m_{REp} V_T} \right) - 1 \right]. \quad (8.92)$$

The ratio  $\gamma_B$  of the perimeter component to its area related counterpart (cf. (5.107)) can be determined experimentally in the same way as  $\gamma_C$  for the collector current. In general,  $\gamma_B \neq \gamma_C$  so that, after scaling the internal component with the effective emitter area (defined by  $\gamma_C$ ), there is either a non-zero remaining contribution given by (8.92) at the perimeter base node B\* or none. In the latter case the total backinjection and recombination current is included in the internal transistor. In other words, in order to keep the same total BE base current the perimeter components, namely  $I_{BEpS}$  and  $I_{REpS}$ , have to be corrected by the amount of current already taken into account by widening the emitter to an effective area. In addition, the internal saturation current values needs to be limited to the total values.

The bias dependence of the perimeter BE depletion charge and capacitance is the same as described in section 8.1.1.1 except that the parameters  $C_{jEp0}$ ,  $V_{DEp}$ ,  $z_{Ep}$  and  $a_{jEp}$  need to be inserted. The actual value of  $C_{jEp0}$ , however, is determined by geometry scaling similarly to  $I_{jBEp}$  described above, since in general the (measurable) ratio  $\gamma_{jE} = C_{jE,P}/C_{jE,A}$  is different from  $\gamma_C$ .

The BE tunneling current is dominated by the region with the highest base doping. While this is the internal transistor in technologies with epitaxial base (SiGe HBTs) it is the perimeter region in implanted transistors or if the spacer width becomes too thin so that the external base doping reaches the emitter doping. In case the tunneling effect is dominated by the

perimeter emitter junction, the variables  $C_{jEi0}$ ,  $V_{DEi}$ ,  $z_{Ei}$ , and  $v_e$  in (8.67) need to be replaced by  $C_{jEp0}$ ,  $V_{DEp}$ ,  $z_{Ep}$  and  $v_{B^*E}/V_{DEp}$ . The resulting tunneling current source  $i_{BEtp}$  in the EC then needs to be connected to the perimeter base node B\* as shown in Fig. 8.15. In order to provide a flexible description within HICUM/L2 and, also, to allow proper geometry scaling, the parameter  $\mathbf{tu}_{\text{node}}$  indicates whether the tunneling current is evaluated for the internal transistor by connection to base node B' ( $\mathbf{tu}_{\text{node}} = 0$ ) or for the perimeter junction by connection to the base node B\* ( $\mathbf{tu}_{\text{node}} = 1$ ).

According to sec. 5.4.6, various isolation capacitances exist between different base and emitter regions (cf. Fig. 5.55). In a compact model, a lumped representation for both the isolation capacitance and the external base resistance is preferred. A  $\pi$ -equivalent circuit is the best first-order compromise for high-frequency modeling without adding nodes. Therefore, the distributed parasitic isolation capacitance

$$C_{BEpar} = C_{Es} + C_{Eo} + C_{Em} \quad (8.93)$$

is partitioned between the perimeter and external base node to make the model more flexible for a larger variety of processes. For instance, if the resistance contribution of the spacer region to the total external base resistance dominates, most of  $C_{BEpar}$  needs to be assigned to the perimeter base node ( $C_{BEpar,2}$ ). The partitioning of  $C_{BEpar}$  in form of a  $\pi$ -equivalent circuit also allows to include the inherent metallization capacitance without any additional effort. The partitioning option is realized by specifying the total parasitic BE capacitance  $C_{BEpar}$  and its partitioning factor  $f_{BEpar}$  as model parameters. The partitioning factor

$$f_{BEpar} = \frac{C_{BEpar,2}}{C_{BEpar}} = \frac{C_{BEiso,2} + C_{BE,metal}}{C_{BEiso} + C_{BE,metal}} \quad (8.94)$$

is defined as the ratio of the “inner” to the total (measured) parasitic capacitance. Thus,  $f_{BEpar} = 0$  indicates that the entire parasitic capacitance is connected between the transistor terminals.

### 8.2.3 External base-collector region

The component for the external BC junction current reads

$$I_{jBCx} = I_{BCxS} \left[ \exp \left( \frac{V_{B^*C'}}{m_{BCx} V_T} \right) - 1 \right]. \quad (8.95)$$

Since the base current across the internal BC junction is scaled with the effective emitter area, the current across the external BC junction is scaled with the total BC junction area minus the effective emitter area.

As discussed in sec. 5.4.5 the total external BC capacitance consists of a bias-dependent external depletion capacitance  $C_{jCx}$ , and a bias-independent parasitic capacitance

$$C_{BCpar} = C_{sti} + C_{Cm} \quad (8.96)$$

that results from the shallow trench isolation ( $C_{sti}$ ) and from metallization ( $C_{Cm}$ ). Again, in order to provide a computationally efficient lumped representation, the *total* capacitance  $C_{BCx} = C_{jCx}(V) + C_{BCpar}$  is partitioned across the external base resistance  $R_{Bx}$ . This corresponds to a first-order high-frequency approximation of the RC transmission line behavior for the external base. As a consequence,  $C_{BCpar}$  and an overall partitioning factor

$$f_{BCpar} = \frac{C_{BCx,2}}{C_{BCx}} \quad (8.97)$$

are introduced as model parameters. However, due to the bias dependence of  $C_{jCx}$  the partitioning has to be performed more carefully as shown in Fig. 8.17. In order to reduce arithmetic operations the *zero-bias* depletion capacitance  $C_{jCx0}$  rather than the voltage dependent depletion capacitance value is partitioned. In the implementation, the connection of the parasitic capacitance between the external nodes and of the depletion capacitances between the inner nodes is given highest priority. Within the model, the capacitances are treated separately due to the bias dependence of  $C_{jCx}$ .

Since the depletion charge of the external BC junction does not depend on the transfer current, the purely voltage dependent expressions given in (8.8) to (8.22) for  $C_{jCi}$  and  $Q_{jCi}$  can be employed by simply inserting the corresponding model parameters and node voltages to obtain  $Q_{jCx1}(C_{jCx01}, V_{DCx}, z_{Cx}, V_{PTCx}, V_{BC'})$  and  $Q_{jCx2}(C_{jCx02}, V_{DCx}, z_{Cx},$

$V_{PTCx}$ ,  $V_{B^*C}$ ), respectively, as well as the associated capacitances  $C_{jCx1}$  and  $C_{jCx2}$ . Finally, the bias-independent parasitic charge components are computed as,

$$Q_{BCpar1} = C_{BCpar1} V_{BC} \quad \text{and} \quad Q_{BCpar2} = C_{BCpar2} V_{B^*C}, \quad (8.98)$$

resulting in the partitioned charges

$$Q_{BCx1} = Q_{BCpar1} + Q_{jCx1}, \quad Q_{BCx2} = Q_{jCx2} + Q_{BCpar2}. \quad (8.99)$$

```

CBCx01 = (1-fBCpar) (CjCx0 + CBCpar)
if(CBCx01 ≥ CBCpar) then           ! connection to B
    CBCpar1 = CBCpar
    CBCpar2 = 0
    CjCx01 = CBCx01 - CBCpar
    CjCx02 = CjCx0 - CjCx01
else                               ! connection to B*
    CBCpar1 = CBCx01
    CBCpar2 = CBCpar - CBCpar1
    CjCx01 = 0
    CjCx02 = CjCx0
endif

```

Fig. 8.17: Implementation of BC depletion and parasitic capacitance partitioning.

### 8.2.4 External series resistances

The resistive regions of the external transistor are represented in Fig. 8.15 by lumped bias independent series resistances. The external base resistance  $R_{Bx}$  and the external collector resistance  $R_{Cx}$ , respectively, are a function of the transistor configuration and design rules as discussed in detail in sec. 5.4.3.2 and sec. 5.4.7. The emitter resistance contributions discussed in sec. 5.4.8 are lumped together into a single resistor  $R_E$ .

### 8.2.5 Collector-substrate junction

The CS depletion charge and capacitance are modelled by the same formulation as for the bottom part of the external BC charge and capacitance, except that the corresponding model parameters  $C_{jS0}$ ,  $V_{jS}$ ,  $z_S$ , and  $V_{PTS}$  have to be inserted. For bulk technologies, the punch-through voltage  $V_{PTS}$  can be set to infinity (i.e. left at its default value). However, for SOI technologies it is necessary to take into account punch-through. Since the CS junction is modelled by a single element,  $C_{jS}$  contains - from a physical point of view - both the bottom and peripheral portion of that junction. According to sec. 5.4.9, the geometry dependence of these components is quite different, depending also on the isolation technique. Thus, proper geometry preprocessing is required to enable using a single equation for accurately describing the total CS depletion capacitance as function of bias.

### 8.2.6 Parasitic substrate transistor

As discussed in sec. 5.4.10 the parasitic substrate transistor can be turned on depending on the transistor structure and layout as well as on the electrical conditions. Following (5.213) the transfer current of the substrate transistor is expressed as,

$$I_{TS} = I_{TSS} \left[ \exp\left(\frac{v_{B^*C}}{m_{Sf}V_T}\right) - \exp\left(\frac{v_{SC}}{m_{Sr}V_T}\right) \right], \quad (8.100)$$

including both forward and inverse operation. In the latter case an additional base current component of the parasitic substrate transistor occurs:

$$i_{jSC} = I_{SCS} \left[ \exp\left(\frac{v_{SC}}{m_{SC}V_T}\right) - 1 \right]. \quad (8.101)$$

Although this current is usually of little *practical* relevance it is generally useful for simulator convergence since it provides a finite junction conductance.

The minority charge storage in the external collector region is taken into account by the most simple formulation for the diffusion charge,

$$Q_{dS} = \tau_{Sf} i_{TSf}. \quad (8.102)$$



Here,  $\tau_{sf}$  represents the average the storage time of the current path under the external base and at the buried layer periphery.

Lateral scaling depends on the substrate-collector structure and the expected dominant effect for turning on the substrate transistor (cf. sec. 5.4.10). Considering the possible structural variants in bipolar technologies special preprocessing appears to be most suitable for geometry scaling.

### 8.2.7 Substrate network

As discussed in sec. 5.4.11, at high frequencies a finite impedance exists between the boundary of the substrate-collector SCR and the substrate contact. The impedance is caused by the ohmic resistance and the permittivity of the substrate as well as possibly by the material properties of the deep trench. Depending on the substrate contact location and the isolation scheme of the transistor the coupling impedance and, hence, the topology (or complexity) of the corresponding coupling network can vary significantly.

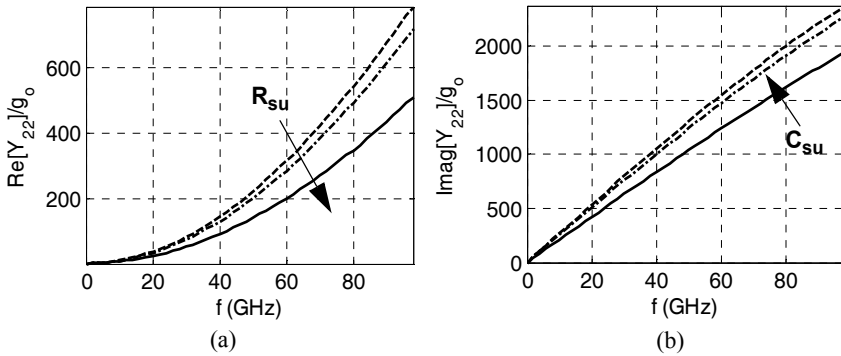


Fig. 8.18: Impact of intra-device substrate-coupling on transistor output admittance, normalized to its zero-bias value  $g_0 = y_{22}(f=0)$ : (a) Real part and (b) imaginary part as function of frequency for different parameter sets:  $T_1$ (--),  $T_2$ (-.), and  $T_3$ (-) with  $\{R_{su}/\Omega, C_{su}/\text{pF}\} = \{T_1(0, 0), T_2(200, 10), T_3(400, 2)\}$ .

For instance, partitioning of the substrate capacitance  $C_{jS}$  into, e.g. an area and a perimeter component, may be required. Furthermore, the length of the current path through the substrate depends on the extension of the SC

SCR into the substrate. However, such a bias dependence can only be taken into account in conjunction with a detailed geometry scaling. The corresponding set of equations is too elaborate for a compact model. As a consequence, HICUM includes only the most simple parallel RC network consisting of a resistance  $R_{Su}$  and a capacitance  $C_{Su}$  as shown in Fig. 8.15 and with the basic physical meaning explained in sec. 5.4.11. For generating an accurate geometry dependent substrate network a special preprocessing appears more suitable.

Figure 8.18 shows the impact of substrate coupling on the frequency dependent output admittance of a transistor for a variation of the model parameters  $R_{Su}$  and  $C_{Su}$ . The case T<sub>1</sub> represents negligible substrate coupling while the other cases take into account a finite substrate impedance.

### 8.3 Temperature dependence

If not specified otherwise, the parameters of a compact model (e.g. in a library) correspond to a given reference temperature  $T_0$  (e.g. 300 K). During simulation, these parameters are then modified for each device according to its actual temperature  $T$ . The latter either is set by the user (as a constant) or varies due to self-heating. As much as possible, the temperature dependence in HICUM/L2 is formulated in terms of the ratio  $T/T_0$ . The corresponding equations are all based on the temperature dependence of physical quantities such as bandgap and mobility (cf. ch. 6).

Compared to the results of ch. 6, some of the original equations have to be modified for numerical reasons (to avoid arithmetic over- or underflow) mostly towards extreme temperatures. The respective smoothing functions used for the simulator implementation are given below. It is assumed that every circuit simulator prevents negative or zero temperature. Since effects such as freeze-out are usually not taken into account by compact models, it is not recommended to use a model at temperatures lower than about 250 K unless its parameters have been extracted or at least been verified especially for that temperature range. As in ch. 6, relative temperature coefficients (TCs) are designated by the symbol  $\alpha$  and absolute TCs by  $\zeta$ .

### 8.3.1 Temperature dependent energy gap or bandgap voltage

In order to allow simulations of devices fabricated in different materials and to make the model simulator-independent, a temperature dependent bandgap voltage is included in HICUM according to

$$V_g(T) = V_g(\mathbf{0}) + f_{2vg}T + f_{1vg}T \ln(T). \quad (8.103)$$

The model parameters  $f_{1vg}$  and  $f_{2vg}$ , respectively, correspond to  $K_1$  and  $K_2$ , respectively, in (6.3) and only depend on the semiconductor material. Since the equation above is applied to all bandgap voltages available in the model, the model parameter  $V_g(\mathbf{0})$  (zero-Kelvin bandgap voltage) needs to take into account average values for possible bandgap changes due to high-doping effects or material composition, making  $V_g(\mathbf{0})$  an effective band-gap voltage.

In advanced BJTs and especially in HBTs the bandgap usually varies significantly in the different transistor regions. Hence, using a single band-gap, typically for the base region and adjusted for the transfer current, can lead to unacceptable errors in the description of other characteristics. Therefore, HICUM allows to specify an average effective bandgap voltage for each transistor region in order to provide the means for very accurate temperature modeling in integrated circuit design. The respective model parameters are  $V_{gE} = V_{gE}(\mathbf{0})$ ,  $V_{gB} = V_{gB}(\mathbf{0})$ ,  $V_{gC} = V_{gC}(\mathbf{0})$ ,  $V_{gS} = V_{gS}(\mathbf{0})$ .

### 8.3.2 Depletion charges and capacitances

The temperature dependent description of the depletion charges is based on the results obtained in sec. 6.3.1.1. The built-in voltage  $V_{Dx}(T)$  for all junctions is modeled in three steps in order to avoid negative values at high temperatures. First an auxiliary voltage is calculated according to (6.24) from the model parameter  $V_{Dx}(T_0)$ ,

$$V_{Dj}(T_0) = 2V_{T0} \ln \left[ \exp\left(\frac{V_{Dx}(T_0)}{2V_{T0}}\right) - \exp\left(-\frac{V_{Dx}(T_0)}{2V_{T0}}\right) \right], \quad (8.104)$$

with  $x = \{Ei, Ep, Ci, Cx, S\}$  and the thermal voltage  $V_{T0} = k_B T_0/q$ . Then, using a temperature dependent average bandgap voltage of the respective junction,

$$V_{g(y,z)} = \frac{V_{gy} + V_{gz}}{2} \text{ with } (y,z) = \{(B,E), (B,C), (C,S)\}, \quad (8.105)$$

the built-in voltage at the actual temperature is calculated from (6.23), (8.104), and the bandgap voltage formulation (8.103), yielding

$$V_{Dj}(T) = V_{Dj}(T_0) \left( \frac{T}{T_0} \right) - m_g V_T \ln \left( \frac{T}{T_0} \right) - V_{g(x,y)} \left( \frac{T}{T_0} - 1 \right). \quad (8.106)$$

For  $m_g = 3$  this reduces to the classical equation, that assumes a linear temperature dependence of  $V_{g(x,y)}$ . Finally, the new built-in voltage is calculated according to (6.20) as

$$V_{Dx}(T) = V_{Dj}(T) + 2 V_T \ln \left( \frac{1}{2} \left[ 1 + \sqrt{1 + 4 \exp \left( -\frac{V_{Dj}(T)}{V_T} \right)} \right] \right) \quad (8.107)$$

All zero-bias capacitance parameters are modeled according to (6.24),

$$C_{jx0}(T) = C_{jx0}(T_0) \left( \frac{V_{Dx}(T_0)}{V_{Dx}(T)} \right)^{z_x}, \quad (8.108)$$

with  $x = \{Ei, Ep, Ci, Cx, S\}$  and the model parameters already given at the reference temperature  $T_0$ .

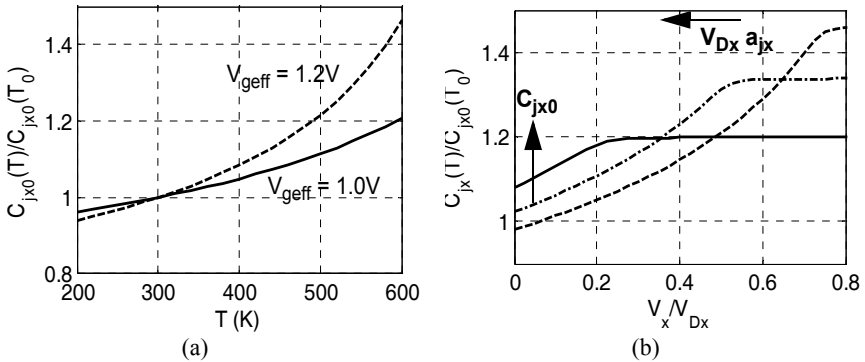


Fig. 8.19: (a) Normalized depletion capacitance versus temperature for different  $V_{geff}$  (b) Voltage dependent normalized depletion capacitance for different temperatures 250K (---), 350K (-), and 450K (· · ·) and the model parameter set  $\{C_{jx0}/\text{fF}, V_{Dx}/V, z_{jx}, a_{jx}\} = \{1, 0.8, 0.33, 1.4\}$  with  $V_{geff} = 1.0V$ .

The parameter  $\alpha_{jx}$  determining the peak value of a depletion capacitance at forward bias decreases with increasing temperature:

$$\alpha_{jx}(T) = \alpha_{jx}(T_0) \frac{V_{Dx}(T)}{V_{Dx}(T_0)}. \quad (8.109)$$

The temperature dependence of the depletion capacitances is illustrated in Fig. 8.19. With increasing temperature, the zero-bias capacitance increases, while the voltage at the peak and the peak itself decrease. Note that neither parasitic capacitances nor the associated partitioning factors  $f_{BEpar}$  and  $f_{BCpar}$  depend on temperature.

### 8.3.3 Transfer current

The temperature dependent transfer current was given by (6.42) and contains the temperature dependent model parameters  $c_{10}$  and  $Q_{p0}$ . Following (6.41), the model equation for the prefactor  $c_{10}$  reads

$$c_{10}(T) = c_{10} \left( \frac{T}{T_0} \right)^{\zeta_{CT}} \exp \left[ \frac{V_{gB}}{V_T} \left( \frac{T}{T_0} - 1 \right) \right] \quad (8.110)$$

with  $\zeta_{CT}$  as model parameter (cf. (6.40) for its physical meaning).

Based on (6.35) the temperature dependence of the zero-bias hole charge  $Q_{p0}$  can be accurately modeled by

$$Q_{p0}(T) = Q_{p0} \left[ (1 + k_{qp0}) - k_{qp0} \left( \frac{V_{DEi}(T)}{V_{DEi}} \right)^{z_{Ei}} \right], \quad (8.111)$$

with  $k_{qp0}$  as model parameter, which is required for SiGe HBTs as compared to  $k_{qp0} = 1$  for BJTs (cf. Fig. 6.2). The temperature modeling of  $V_{DEi}(T)$  ensures that  $Q_{p0}(T)$  will always remain positive without any additional smoothing effort. Above expression (8.111) is preferred over (6.36), although during parameter extraction inaccuracies of  $Q_{p0}(T)$  (and also of the depletion charges) will be at least partially compensated through  $\zeta_{CT}$ . However,  $\zeta_{qp0}$  in (6.36) would still be required as additional parameter, since  $Q_{p0}(T)$  also enters the charge sum in the transfer current expression. Hence, if a new model parameter needs to be introduced in any way it makes more sense to use the more accurate expression (6.35).

### 8.3.4 Transit time and minority charge

An important variable for modeling of the minority charge is the critical current  $I_{CK}$  as defined in (8.27). It depends on temperature via the parameters  $r_{Ci0}$ ,  $V_{lim}$ , and  $V_{CEs}$ , which in turn depend on physical parameters like mobility of the collector and saturation velocity. The temperature dependence of the latter variables was given in sec. 6.2.

The internal collector resistance contains the low-field electron mobility as the only temperature dependent variable and reads after inserting (6.11) into (3.243)

$$r_{Ci0}(T) = r_{Ci0} \left( \frac{T}{T_0} \right)^{\zeta_{Ci}}. \quad (8.112)$$

The parameter  $\zeta_{Ci}$  can be estimated as function of collector doping from (6.12).

The voltage  $V_{lim}$  was defined in (3.10) and contains both collector mobility and saturation velocity (through the critical field  $E_{lim}$ ). Inserting (6.10) and (6.12) results in

$$V_{lim}(T) = V_{lim} \left( \frac{T}{T_0} \right)^{\zeta_{Ci} - \zeta_{vs}} \quad (8.113)$$

which is simple and numerically stable, and does not require any additional model parameters than those of the physical variables. The equation is valid up to about 600K, which is the highest temperature of available experimental data for mobility and saturation velocity in various semiconductor materials.

The CE saturation voltage is modelled as linear function of temperature,

$$V_{CEs}(T) = V_{CEs} [1 + \alpha_{CEs}(T - T_0)], \quad (8.114)$$

with the relative TC  $\alpha_{CEs}$  as model parameter, the physical meaning of which is given by the difference of the TCs of  $V_{DEi}$  and  $V_{DCi}$ .

The following discussion on the transit time is based on the results of sec. 6.3.1.2. The low-current portion of the transit time,  $\tau_{f0}$ , as a function of temperature is mainly determined by the quadratic temperature dependence of the parameter

$$\tau_0(T) = \tau_0 [1 + \alpha_{\tau_0}(T - T_0) + k_{\tau_0}(T - T_0)^2] \quad (8.115)$$

with the relative TCs  $\alpha_{t0}$  and  $k_{t0}$  as model parameters that can be related to physical variables [3].

The base and collector component of  $\tau_f$  depend on temperature via the same diffusivity (of the collector). Therefore, the temperature dependence of the composite parameter  $\tau_{hcs}$  can be expressed as

$$\tau_{hcs}(T) = \tau_{hcs}\left(\frac{T}{T_0}\right)^{(\zeta_{ci}-1)}. \quad (8.116)$$

The emitter time constant  $\tau_{Efo}$  depends on temperature according to (6.27). Assuming a high emitter concentration gives  $a_{tfe0} \approx 0.5$ . Furthermore, inserting  $a_{Bfi} = \zeta_{CT} - \zeta_{BET}$  yields

$$\tau_{Efo}(T) \cong \tau_{Efo}\left(\frac{T}{T_0}\right)^{\zeta_{BET} - \zeta_{CT} - 0.5} \exp\left[\frac{V_{gB} - V_{gE}}{V_{T0}}\left(1 - \frac{T_0}{T}\right)\right] \quad (8.117)$$

that is also numerically stable for all temperatures.

The parameter  $\tau_{Bfvs}$  used in (8.46) is defined in (3.278) and depends on temperature via the low-field collector diffusivity and the base drift factor. Making the reasonable assumption for advanced transistors that  $\zeta_h \geq 2$  so that  $f\zeta_h \gg 1$  and, hence,  $G\zeta_h \approx \zeta_h$  leads to the expression

$$\tau_{Bfvs}(T) = \tau_{Bfvs}\left(\frac{T}{T_0}\right)^{\zeta_{ci}}. \quad (8.118)$$

The collector barrier voltage  $V_{Cbar}$  can be described according to (6.28) by setting  $\Delta V_{Cb} = \Delta V_V/2$ . However, considering the fact that the influence of the barrier only occurs at high current densities when self-heating increases the device temperature significantly, the  $\ln(\cdot)$  term in (6.28) has only a minor influence in the temperature range considered here. Therefore, the temperature dependence of the barrier voltage parameter in (4.8) can be approximated as

$$V_{Cbar}(T) = V_{Cbar}[1 + \alpha_{Cbar}(T - T_0)] \quad (8.119)$$

with its relative TC

$$\alpha_{Cbar} = \frac{a_{V2}}{2T_0 V_{Cbar}} \quad (8.120)$$

and  $a_{V2}$  from (6.29). Since a calculation of  $a_{V2}$  and thus  $\alpha_{Cbar}$  from physical parameters and variables requires to know the material composition  $C$

and the dependence of the valence band change with  $C$  as well as the TC of the electron affinity, it is more convenient from an application point of view to define  $\alpha_{Cbar}$  as model parameter.

The temperature dependence of the various minority charge components and of the NQS delay times follows automatically from that of the transit time.

### 8.3.5 Junction current components

The temperature dependent saturation current of the BE backinjection component is given by (6.49),

$$I_{BEiS}(T) = I_{BEiS}\left(\frac{T}{T_0}\right)^{\zeta_{BET}} \exp\left[\frac{V_{gE}}{V_T}\left(\frac{T}{T_0} - 1\right)\right], \quad (8.121)$$

with the model parameters  $\zeta_{BET}$  and the effective emitter bandgap voltage  $V_{gE}$ . An estimate for  $V_{gE}$  can be calculated from the effective bandgap voltage in the base and the measured relative TC  $\alpha_{Bf}$  of the current gain:

$$V_{gE} = V_{gB} - \alpha_{Bf} T_0 V_{T0}. \quad (8.122)$$

This estimate can be used as default or preliminary value, e.g., when converting the parameter  $\alpha_{Bf}$  of older HICUM/L2 versions (up to v2.21) to  $V_{gE}$ . The reasons for not using the current gain as variable have been discussed in sec. 6.3.2.5.

The BE recombination component is given by (6.50)

$$I_{REiS}(T) = I_{REiS}\left(\frac{T}{T_0}\right)^{\zeta_{REi}} \exp\left[\frac{V_{gBE}}{2V_T}\left(\frac{T}{T_0} - 1\right)\right], \quad (8.123)$$

with  $\zeta_{REi} = m_g/2$  and  $m_g = 3 - f_{Ivg}q/k_B$  from (6.9) as well as the average bandgap voltage  $V_{gBE}$  defined by (8.105). Obviously, (8.123) does not require any additional parameters. Assuming the recombination time constant  $\tau_{Bhrec}$  to be temperature independent, the component  $I_{Bhrec}$  becomes automatically temperature dependent via the excess hole charge  $\Delta Q_{Bf}$ .

All other junction related saturation currents are also described by formulations similar to (8.121) or (8.123) but with the parameters listed in Table 8.1. The bandgap voltages as well as the factors  $\zeta_{Ci}$  and  $\zeta_{Cx}$  have already been introduced earlier. The only new model parameter is  $\zeta_{BET}$  since all other temperature factors  $\zeta$  can be expressed by already existing



parameters. Due to the low substrate doping, the mobility exponent factor  $\zeta_{mpS} = 2.5$  can be assumed as a good approximation. This approach allows to save a number of model parameters that exist in some other models but which do not significantly increase accuracy and flexibility, since they are of *little importance* for *circuit design*. This also keeps the model as simple as possible for parameter determination.

component	eq.	bandgap $V_{geff}$	factor $\zeta$
$I_{BEiS}$	(8.121)	$V_{gE}$	$\zeta_{BET}$
$I_{BEpS}$	(8.121)	$V_{gE}$	$\zeta_{BET}$
$I_{REiS}$	(8.123)	$V_{gBE}$	$\zeta_{REi} = m_g/2$
$I_{REpS}$	(8.123)	$V_{gBE}$	$\zeta_{REp} = m_g/2$
$I_{BCiS}$	(8.121)	$V_{gC}$	$\zeta_{BCi} = m_g + 1 - \zeta_{Ci}$
$I_{BCxS}$	(8.121)	$V_{gC}$	$\zeta_{BCxT} = m_g + 1 - \zeta_{Cx}$
$I_{SCS}$	(8.121)	$V_{gS}$	$\zeta_{SCT} = m_g + 1 - \zeta_{\mu pS}$

Table 8.1: Junction saturation current components and their parameters for modeling the temperature dependence. The column “eq.” indicates the formulation employed. For silicon, the values of the constants are:  $m_g = 4.188$ ,  $\zeta_{\mu pS} = 2.5$ .

The temperature dependent local avalanche current in (8.66) is given by the temperature dependence of the transfer current, internal BC depletion capacitance, and the model parameters  $f_{AVL}$  and  $q_{AVL}$ . The latter are given by (8.65). Inserting the factors  $a_n$  and  $b_n$  from (6.13a) yields

$$f_{AVL}(T) = f_{AVL}(T_0) \frac{1 - \alpha_{an} \Delta T}{1 - \alpha_{bn} \Delta T}, \quad (8.124)$$

$$q_{AVL}(T) = q_{AVL}(T_0) [1 - \alpha_{bn} \Delta T], \quad (8.125)$$

with  $\alpha_{an}$  and  $\alpha_{bn}$  as model parameters. Their default values can be taken directly from the literature. The temperature dependence of the non-local avalanche current, as given by (6.51), requires the same model parameters.

The temperature dependence of the tunneling current (8.67) is determined by its saturation current and the exponent coefficient (8.68). Writing the general form for the internal or perimeter junction gives

$$I_{btbS}(T) = I_{btbS} \sqrt{\frac{V_{gBE}(T_0)}{V_{gBE}(T)}} \left( \frac{V_{DE}(T)}{V_{DE}} \right)^2 \frac{C_{jE0}(T)}{C_{jE0}}, \quad (8.126)$$

$$a_{btb}(T) = a_{btb} \left( \frac{V_{gBE}(T)}{V_{gBE}(T_0)} \right)^{3/2} \frac{V_{DE}}{V_{DE}(T)} \frac{C_{jE0}}{C_{jE0}(T)}. \quad (8.127)$$

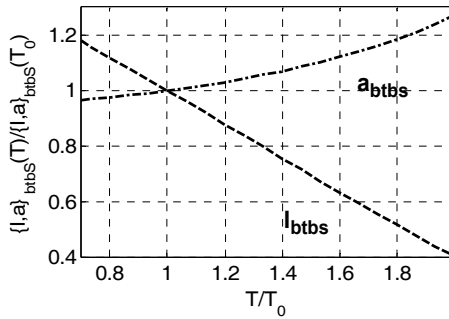


Fig. 8.20: Normalized tunneling saturation current (--) and exponent coefficient (--) vs. normalized temperature.

In contrast to (8.105), here the average bandgap voltage of base and emitter has to be calculated *at each temperature*:

$$V_{gBE}(T) = \frac{V_{gB}(T) + V_{gE}(T)}{2}. \quad (8.128)$$

Furthermore, for  $(V_{DE}, C_{jE0})$  the parameters for the internal junction  $(V_{DEi}, C_{jEi0})$  or the perimeter junction  $(V_{DEp}, C_{jEp0})$  have to be inserted depending on which tunneling current component is to be evaluated. The temperature dependence of the normalized variables of  $I_{btbS}$  and  $a_{btb}$  is illustrated in Fig. 8.20.

### 8.3.6 Series resistances

The temperature dependence of all series resistances is in general determined by the mobility as well as contact or interface recombination velocity of the respective regions. According to (6.55) the zero-bias internal base resistance reads

$$r_{Bi0}(T) = r_{Bi0} \left( \frac{T}{T_0} \right)^{\zeta_{rBi}}. \quad (8.129)$$

The model parameter  $\zeta_{rBi}$  accounts for the (average) base doping concentration according to (6.12). Conductivity modulation and emitter current crowding in  $R_{Bi}$  are automatically described as a function of  $T$  by the corresponding charges and currents.

The external base resistance  $R_{Bx}$ , external collector resistance  $R_{Cx}$ , and emitter series resistance  $R_E$  follow the formulation (8.129) with corresponding model parameters  $\zeta_{rBx}$ ,  $\zeta_{rCx}$  and  $\zeta_{rE}$  that include the effect of both (average) doping concentrations within the corresponding regions and contact or interface recombination velocity.

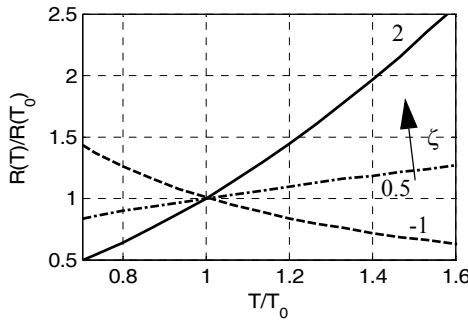


Fig. 8.21: Normalized resistance as a function of normalized temperature according to (8.129) for different values of  $\zeta$  ( $= \zeta_{Ci}, \zeta_{rBi}, \zeta_{rBx}, \zeta_{rCx}$  or  $\zeta_{rE}$ ) as parameter: -1 (---), 0.5 (---), and +2 (-).

Figure 8.21 shows the types of temperature dependence that can be modelled with (8.129). The temperature dependence of the substrate coupling resistance  $R_{su}$  can easily be included employing (8.129) but with  $\zeta_{mpS} = 2.5$ .

Note that the internal base shunt capacitance  $C_{rBi}$  (parallel to  $R_{Bi}$ ) is temperature dependent via the capacitances of the internal transistor.

### 8.3.7 Parasitic substrate transistor

The substrate transistor storage time depends on temperature via the diffusivity in the external collector region,

$$\tau_{sf}(T) = \tau_{sf}\left(\frac{T}{T_0}\right)^{(\zeta_{Cx}-1)}, \quad (8.130)$$

with  $\zeta_{Cx}$  as a model parameter that can be estimated from (6.12) as function of external collector doping.

The substrate transistor transfer current (8.100) contains a temperature dependent saturation current that can be described analogously to (8.121)

$$I_{TSS}(T) = I_{TSS}\left(\frac{T}{T_0}\right)^{\zeta_{BCxT}} \exp\left[\frac{V_{gC}}{V_T}\left(\frac{T}{T_0} - 1\right)\right] \quad (8.131)$$

with the factor  $\zeta_{BCxT}$  from Table 8.1.

## 8.4 Self-Heating model

As elaborated in sec. 6.4.1, the description of self-heating can be transformed into an electrical equivalent circuit. The most simple form for calculating the increase  $\Delta T_j$  of the junction temperature is the thermal network shown in Fig. 8.22. The current source corresponds to the power dissipated in the device, and the node voltage corresponds to  $\Delta T_j$ . Hence, under DC conditions  $T = T_j = T_0 + \Delta T_j$  is fed back to the calculation of the elements of the transistor EC. This calculation requires the thermal resistance  $R_{th}$  of the particular transistor structure as model parameter. Such a simple representation of self-heating is sufficient for the vast majority of DC applications if  $R_{th}$  is properly determined.

For time varying signals the power dissipation and, thus, the heat flow within a device also varies with time. Therefore, the device temperature, e.g. at the BE junction, changes with time. The corresponding time constants of this junction temperature variation are determined by the material in the path of the heat flow. The specific heat capacitance of the different

materials leads to a delayed reaction of the temperature change with power dissipation change. This delay, again in its most simple form, can be represented in the thermal EC by a thermal capacitance  $C_{th}$ , resulting in the single-pole thermal network shown in Fig. 8.22 that is also used in HICUM as built-in network. This thermal network is solved together with the electrical EC of a transistor if  $R_{th} > 0$  and the self-heating flag  $f_{SH}$  is turned on.

Note that only *self*-heating can be taken into account by the network in Fig. 8.22 but not the thermal coupling between different devices on the chip as discussed in sec. 6.4.2. In order to allow more accurate modeling of self-heating, e.g. by multi-pole thermal networks, or inter-device thermal coupling, the temperature node  $\Delta T_j$  in Fig. 8.22 has been made accessible from the outside. In this case, the impact of the internal elements can be eliminated by setting  $R_{th}$  to a very large value and  $C_{th} = 0$ .

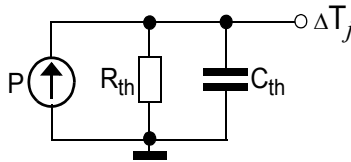


Fig. 8.22: Adjunct self-heating network associated with the complete HICUM model in Fig. 8.15.

Generally, the power is calculated from all relevant dissipative elements in the equivalent circuit and thus reads

$$P = |I_T V_{C'E}| + |I_{AVL} V_B \cdot C| + \sum |\Delta V_n|^2 / R_n + |I_{TS} V_B \cdot S| \quad (8.132)$$

with  $R_n = \{R_{Bi}, R_{Bx}, R_E, R_{Cx}\}$  as non-zero series resistances and  $\Delta V_n$  as the corresponding voltage drop across those resistances. Contributions from current across junctions,  $I_j(V_D - V)$ , are negligible at low injection due to low  $I_j$  and at high injection due to the vanishing potential drop  $(V_D - V)$ . (8.132) is exercised for  $f_{SH} = 2$ .

However, since the consideration of *all* dissipative elements generates elaborate expressions for the derivatives in the Jacobian, which can lead to significant computational effort, setting  $f_{SH} = 1$  only includes the most relevant dissipative elements in the power calculation

$$P = |I_T V_{C'E}| + |I_{AVL} V_B \cdot C|. \quad (8.133)$$

This simplification, which was used in the earlier model versions, worked well and seems justified in many cases since the accuracy of not only the single-pole network itself but also of the values of  $R_{th}$  and  $C_{th}$  can be fairly limited. Figure 8.23 shows the effect of self heating on collector current for different values of  $\beta_{SH}$ . In this case, there is essentially no difference between using (8.132) and (8.133). However, the model with (8.132) requires about two to three more iterations to converge compared to that with (8.133). The number of iterations depends upon the given bias point and the values of  $R_{th}$ . With  $R_{th} = 5000$  K/W, the number of iterations varies between 14 ( $\beta_{SH} = 1$ ) and 16 ( $\beta_{SH} = 2$ ).

For AC and transient simulation, also  $C_{th}$  must have a realistic value. Otherwise the y-parameters and, for instance, the determination of the transit frequency becomes incorrect. Keeping  $C_{th} = 0$  during these operations is an often found mistake made by model users.

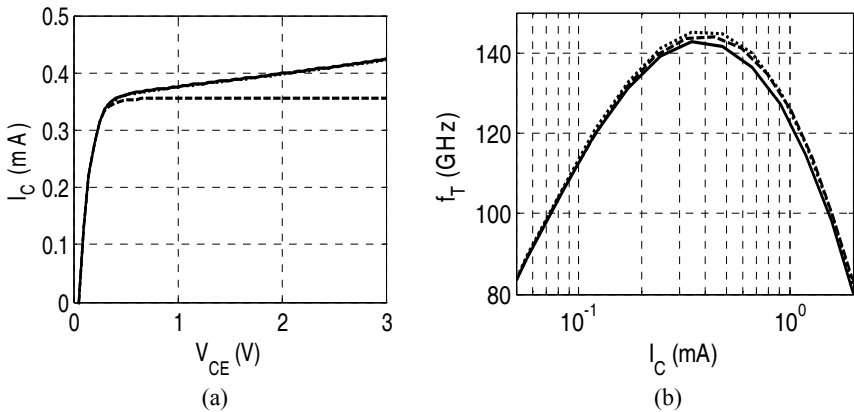


Fig. 8.23: Effect of self-heating at 300K: (a) Output characteristics at  $V_{BE}=0.9$ V for  $\beta_{SH} = 0$  (dashed line),  $= 1$  (dotted line), and  $= 2$  (solid line) with  $(R_{TH}, C_{TH}) = (5000 \text{ K/W}, 0.5\text{E-}9\text{Ws/K})$ . (b) Transit frequency with  $(R_{TH}, C_{TH}) = (5000 \text{ K/W}, 0)$  for  $\beta_{SH} = 0$  (dashed line) and  $= 1$  (dotted line), and with  $(R_{TH}, C_{TH}) = (5000 \text{ K/W}, 0.5\text{E-}9\text{Ws/K})$  for  $\beta_{SH} = 1$  (solid line).

According to (6.58) the thermal conductivity depends on temperature. For not too large temperature changes the relatively weak dependence, at least in silicon based transistors, can often be neglected since the error is smaller than the uncertainty about the thermal resistance value. However, this may not be possible anymore for larger temperature changes and some

other materials. A suitable procedure for calculating the thermal resistance as a function of junction temperature from the given parameter reference temperature  $T_0$  and the thermal resistance parameter  $R_{th}(T_0)$  was described in sec. 6.4.1. Combining (6.63) and (6.65) yields

$$R_{th}(P, T_j) = R_{th} \left[ 1 + \alpha_{th, T} \left( \frac{T_j}{T_0} - 1 \right) \right] [1 + \alpha_{th, P} P] \quad (8.134)$$

with the model parameters  $\alpha_{th, T}$  and  $\alpha_{th, P}$ . The default value of  $\alpha_{th, T}$  is given in Table 6.2, while  $\alpha_{th, P}$  is given by (6.65).

## 8.5 Small-signal model

The exact small-signal EC can be derived from the large-signal EC in Fig. 8.15. This is in fact done automatically in the actual model implementation using Verilog-A code and a model compiler. The latter generates all existing derivatives of charges and currents as well as the associated elements in the small-signal EC. Since the resulting overall small-signal EC is quite elaborate, it will not be presented here and neither will be the equations of its elements. However, as an aid for circuit design, the simplified small-signal EC shown in Fig. 8.24 consisting of the most important elements will be briefly discussed below for the most relevant applications at forward operation, i.e. at  $V_{C'E'} > 2V_{CES}$ .

The EC of the internal transistor in Fig. 8.24a includes the transfer current source  $I_T$  described by the transconductance

$$g_m \cong \left. \frac{dI_{Tf}}{dV_{B'E'}} \right|_{V_{CE}} \quad (8.135)$$

and the corresponding (forward) output conductance

$$g_o \cong \left. \frac{dI_{Tf}}{dV_{C'E'}} \right|_{V_{B'E'}} \quad (8.136)$$

If non-quasi-static effects are included,  $g_m$  becomes a complex variable. The internal base current components are represented by their conductances

$$g_{BEi} = \left. \frac{d(I_{jBEi} + I_{Bhrec})}{dV_{B'E'}} \right|_{V_{B'C}} , \quad g_{jBCi} = \frac{dI_{jBCi}}{dV_{B'C}} . \quad (8.137)$$

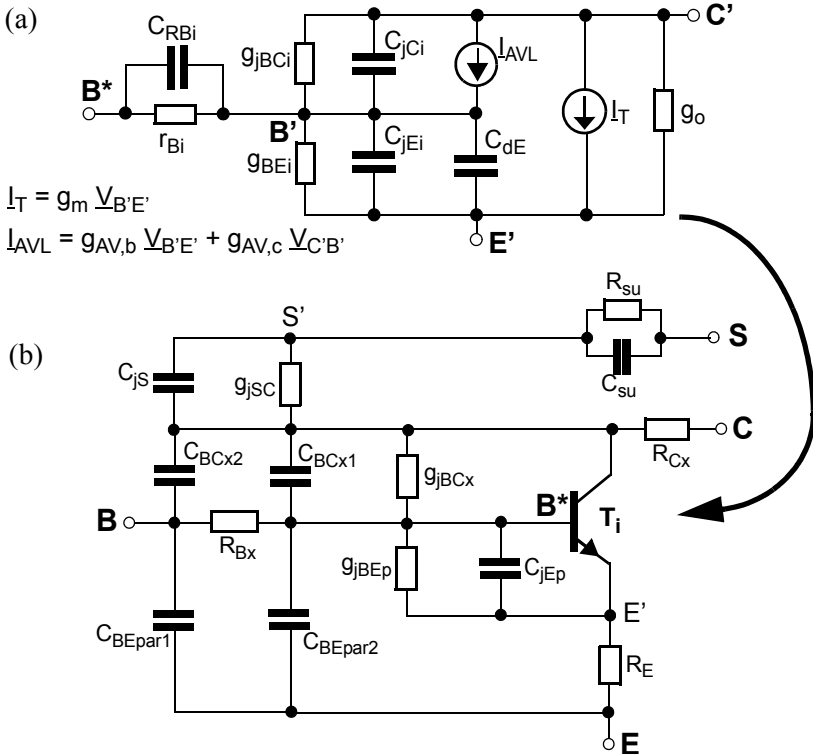


Fig. 8.24: Simplified small-signal EC of HICUM/L2 including the most important elements at forward operation ( $V_{C'E'} > 2 V_{CEs}$ ) for (a) internal and (b) complete transistor. Elements resulting from self-heating are not included in this example.

The (trans)conductances of the avalanche current follow directly from (8.66) as

$$g_{AV,b} = \left. \frac{dI_{AVL}}{dV_{B'E'}} \right|_{V_{B'C}} , \quad g_{AV,c} = \left. \frac{dI_{AVL}}{dV_{B'C}} \right|_{V_{B'E'}} . \quad (8.138)$$

The internal depletion capacitances  $C_{jEi}$  and  $C_{jCi}$  were given in section 8.1.1, while the BE diffusion capacitance is given by



$$C_{dE} = \tau_f g_m. \quad (8.139)$$

The small-signal internal base resistance reads

$$r_{Bi} = \frac{R_{Bi}}{1 - I_{Bi} \frac{dR_{Bi}}{dV_{B'}}} \approx R_{Bi} \left[ 1 + I_{Bi} \frac{dR_{Bi}}{dV_{B'}} \right] \quad (8.140)$$

with  $I_{Bi} = I_{jBEi} + I_{jBCi} + I_{AVL} + I_{Bhrec}$  as total DC internal base current. Finally, dynamic small-signal emitter current crowding is accurately modeled by (8.76).

The EC of the external transistor in Fig. 8.24b includes the conductances  $g_{jBEp}$ ,  $g_{jBCx}$ , and  $g_{jSC}$  of the BE, BC and substrate junction currents, respectively, that are defined similarly to (8.137). The corresponding depletion capacitances  $C_{jEp}$ ,  $C_{jBCx}$ , and  $C_{jS}$ , were defined by the equations in section 8.1.1. The remaining elements, including the external series resistances, the parasitic capacitances, and the substrate coupling network, are bias independent and, hence, maintain the same values as for the DC case.

If self-heating is turned on, the derivatives of the relevant equivalent circuit element variables in Fig. 8.24 with respect to the temperature caused by the dissipated power  $P$  also need to be taken into account, leading to additional controlled sources driven by  $\Delta T_j$ .

## 8.6 Noise model

The physical origin of noise in BJTs and HBTs was discussed in ch. 7. Below, the results will be summarized in terms of the relevant equivalent noise currents  $I_{X,n} = \sqrt{S_{I_X}(f)}$  for small-signal frequency dependent simulation; here, X indicates the corresponding component (e.g., X = RBi) and n indicates "noise". The current-based formulation is preferred since it does not add undesired nodes to the existing small-signal equivalent circuit. Figure 8.25 shows the simplified small-signal noise EC of HICUM/L2 with the relevant noise sources. As discussed in section 8.5 the complete small-signal EC generated automatically by the Verilog-A model compiler is more sophisticated.

All series resistances generate thermal noise as described by (7.27),

$$I_{R,n} = \sqrt{\frac{4k_B T \Delta f}{R}}, \quad (8.141)$$

where  $R = \{R_E, R_{Cx}, R_{Bx}, r_{Bi}\}$ ,  $T$  is the absolute device temperature, and  $\Delta f$  is the frequency interval. Note that the small-signal value  $r_{Bi}$  should be inserted and that there should be no noise assigned to the thermal resistance  $R_{th}$ .

All diode currents injected across junctions are described according to (7.75) but with the corresponding quadratic frequency term neglected,

$$I_{X,n} = \sqrt{2qI_X \Delta f}, \quad (8.142)$$

and with  $X = \{BEi, BCi, BEp, BCx, SC\}$ . Above expression corresponds to pure shot noise. According to (7.70) this is also assumed for the substrate transfer current  $I_{TS}$  (i.e.  $X = TS$  in (8.142)). Noise contributions from recombination current components are presently neglected. This includes  $I_{Bhrec}$ , which becomes relevant only at very high current densities where noise is of little importance for practical applications. Generally though recombination and random telegraph noise could be easily added through a separate Lorentzian noise term [9].

Shot noise is also generated by the avalanche current  $I_{AVL}$ , but needs to be described differently. For sufficiently weak impact ionization (7.78) can be used, resulting in

$$I_{AVL,n} = \sqrt{2qMI_{AVL} \Delta f} \quad (8.143)$$

with  $M = 1 + I_{AVL}/I_T$  and  $I_{AVL}$  from (8.66).

The base current components injected across the BE junction also contain flicker noise (cf. sec. 7.1.5), which depends inversely on the frequency  $f$ . Investigations of flicker noise in polysilicon-emitter bipolar transistors seem to indicate that flicker noise is generated mainly at the poly-silicon to mono-silicon interface. This corresponds to a strong correlation between the bottom and perimeter component. As a consequence, and for simplification of the noise model and its implementation, the present version of HICUM/L2 combines the bottom and perimeter current into the single flicker noise source

$$I_{BE x, F} = \sqrt{k_F (I_{jBEi} + I_{jBEp})^{a_F} \frac{\Delta f}{f}} \quad (8.144)$$

with  $x = \{i, p\}$ . Depending upon the value of a flag  $\mathbf{cf}_{BE}$ , above flicker noise source is connected to either the internal base node ( $\mathbf{cf}_{BE} = -1$ ) or the perimeter base node ( $\mathbf{cf}_{BE} = -2$ ). The value range  $[0,1]$  is reserved for a correlation factor in a possible future development and implementation.

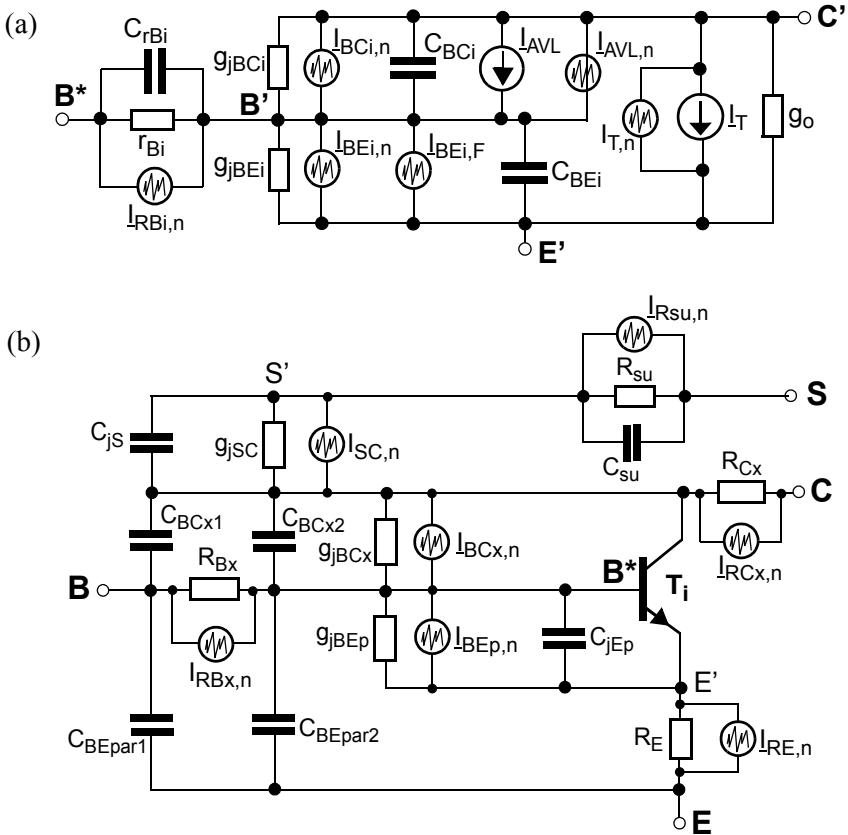


Fig. 8.25: Noise equivalent circuit of HICUM/L2 (simplified and without correlated noise): (a) internal transistor ( $C_{BEi} = C_{jEi} + C_{dE}$  and  $C_{BCi} = C_{jCi} + C_{dC}$ ); (b) complete transistor.

Fluctuations related to the main transfer current  $I_T$  are given for sufficiently low frequencies by (7.70) as pure shot noise:

$$I_{T,n} = \sqrt{2qI_T\Delta f}. \quad (8.145)$$

However, for high frequencies the correlation between the base and transfer current shot noise reduces the overall noise. The importance for taking into account this correlation is increasing in advanced HBTs as was shown in [10, 11]. The results of sec. 7.2.1 to 7.2.3 can be converted into the EC shown in Fig. 8.26a. It contains the independent noise source for BE back-injection ( $I_{BEi,n}$ ) as well as the correlated noise sources for the transfer current ( $I_{T,n}$ ) and for the current associated with diffusion charge in the neutral base ( $I_{Bb,n}$ ). Note, that  $I_{Bb,n}$  and  $I_{BEi,n}$  are completely uncorrelated.

The practical realization of the correlation in a circuit simulator can be done in different ways depending on the noise analysis algorithm and the transformation of the noise sources in the final EC. A possible solution was presented in [12]. There, the noise current  $I_{Bb,n}$  was partitioned into a component  $I_{Bb,nu}$  that is uncorrelated to  $I_{T,n}$  and a component  $I_{Bb,nc}$  that is fully correlated to  $I_{T,n}$ . A serious limitation is that SPICE-like simulators treat all noise sources as *independent* and do not allow to include correlation directly in the analysis. Hence,  $I_{Bb,nc}$  then would need to be transformed to the output and combined with  $I_{T,n}$ . The problem is further aggravated in Verilog-A where shot or white noise expressions cannot be frequency dependent and negative. A possible solution of this problem was presented in [11] employing a system-theoretical approach in which the correlated noise sources were transformed into sources that can be handled like uncorrelated sources by the circuit simulator and lead to the EC in Fig. 8.26b. Here,  $I_{Bb,nu}$  has already been neglected due to the quadratic frequency dependence of its power spectral density (PSD) compared to the frequency independent PSD of  $I_{BEi,n}$ , so that the PSD of the uncorrelated portion of the combined base current  $I_{BE} = I_{BEi} + I_{Bb}$  reads

$$S_{I_b}(\omega) = S_{I_{BEi}} + S_{I_{Bb,u}} \cong 2qI_{BEi}. \quad (8.146)$$

The complex valued correlation factor  $\underline{c}_{BT}$  is then defined for the combined base current and follows from (7.69) and, more generally, (7.72):

$$\underline{c}_{BT} = \frac{S_{I_{Bb},I_T}}{S_{I_b}} \cong -j\omega \left( \frac{\tau_{Bfd}}{3} + \tau_{BC} \right) \frac{I_T}{I_{BEi}}. \quad (8.147)$$

The time constant  $\tau_{Bfd}/3 + \tau_{BC}$  corresponds to the NQS effect related delay time  $\tau_{IT}$  for a diffusion transistor at low current densities. Generalizing this time constant, as was done in (8.61), and defining the intrinsic current gain  $B_{fi} = I_T/I_{BEi}$  yields

$$\mathcal{C}_{BT} = -j\omega\tau_{IT}B_{fi}. \quad (8.148)$$

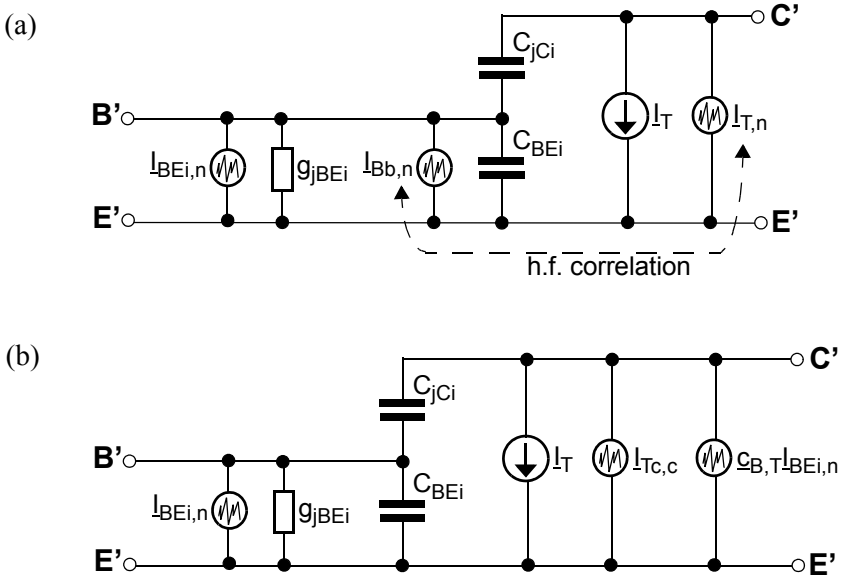


Fig. 8.26: Noise EC for handling the correlation between base and transfer current: (a) EC with physical noise sources; (b) EC with transformed noise sources for circuit simulation.

Furthermore, the transformation gives for the modified transfer current noise source

$$S_{I_{Tc}}(\omega) = S_{I_T}[1 - B_{fi}(\omega\tau_{IT})^2] \quad (8.149)$$

with  $S_{I_T} = 2qI_T$ .

The negative sign in  $\mathcal{C}_{BT}$  makes the realization of the correlation in Verilog-A difficult to implement since the calculations in SPICE-like simulators are not based on the noise PSD, but the latter are obtained as a squared quantity, leading always to a cancellation of the negative sign. Hence, the following approximation is used,

$$S_{I_{Tc}} = S_{I_T} [1 - B_{fi}(\omega\tau_{IT})^2]^2 = S_{I_T} |\mathcal{C}_{BT}|^2 S_{I_b} + \underbrace{\frac{|\mathcal{C}_{BT}|^4}{4B_{fi}} S_{I_b}}_{\text{error term}},$$

in which last term causes an error at very high frequencies. Note that the PSD of the noise current flowing into the internal collector node C' in Fig. 8.26 is *not*  $S_{I_{Tc}} + (\mathcal{C}_{BT} I_{BEi,n})^2 = 2qI_T$  since the noise source  $\mathcal{C}_{BT} I_{BEi,n}$  only yields a contribution *together* with the source  $I_{BEi,n}$ ; it is *zero* if  $I_{BEi,n}$  is not “turned on” in the calculation.

The noise sources described so far are calculated from DC currents and are evaluated in a circuit simulator during a small-signal analysis in frequency domain. However, as pointed out in [9] the impact of noise during nonlinear device operation can be important. In this case, the time dependent currents need to be employed for calculating the noise (current or voltage) sources. While a Verilog-A based implementation with up to three adjunct networks per time dependent nonlinear noise source is possible [9] it is preferred to rather automate this capability as well as the incorporation of correlations in circuit simulators in order to provide a direct and computationally more efficient solution.

## 8.7 Model parameter list

Table 8.2 summarizes the HICUM/L2 model parameters with their brief description, allowed value range, and unit.

The list is divided into groups of parameters according to the elements in the HICUM/L2 equivalent circuit shown in Figs. 8.2 and 8.15 as well as those for additional physical effects such as temperature dependence and noise. As typical for advanced compact models, the total number of model parameters appears to be large due to the many independent physical effects (and their flags) that are being taken into account. However, the modularity of the model formulation and a clearly defined extraction procedure (implemented in existing commercial infrastructures) allow to determine the relevant parameters step by step. “Relevant” means that for given applications and a given process technology not always *all* parameters have to be extracted. For example, certain parameters are related to HBTs only

and, therefore, can be left at their default values for homojunction transistors.

```

simulator lang = spectre
subckt hicumL2_model_default c b e s qhcm (c b e s) hic2_full
+ c10 = 2.0E-30 qp0 = 2.0E-14 ich = 0.0 mcf = 1.0 hfe = 1.0 hfc = 1.0
+ hjei = 1.0 hjci = 1.0 ibeis = 1.0E-18 mbei = 1.0 ireis = 0.0 mrei = 2.0
+ ibeps = 0.0 mbep = 1.0 ireps = 0.0 mrep = 2.0 tbhrec = 0.0
+ ibcis = 1.0E-16 mbci = 1.0 ibcx = 0.0 mbcx = 1.0 ibets = 0.0
+ abet = 40 tunode = 1 favl = 0.0 qavl = 0.0 alfav = 0.0 alqav = 0.0
+ rbi0 = 0.0 rbx = 0.0 fgeo = 0.6557 fdqr0 = 0.0 fcrbi = 0.0 fqi = 1.0
+ re = 0.0 rcx = 0.0 +itss = 0.0 msf = 1.0 iscs = 0.0 msc = 1.0 tsf = 0.0
+ rsu = 0.0 csu = 0.0 cjei0 = 1E-20 vdei = 0.9 zei = 0.5 ajei = 2.5
+ cjep0 = 1E-20 vdep = 0.9 zep = 0.5 ajep = 2.5 cjci0 = 1E-20
+ vdci = 0.7 zci = 0.4 vptci = 100 cjcx0 = 1E-20 vdcx = 0.7 zcx = 0.4
+ vptcx = 100 fbc = 0.0 fbepar = 1.0 cjs0 = 0.0 +vds = 0.6 zs = 0.5
+ vpts = 100 t0 = 0.0 dt0h = 0.0 tbvl = 0.0 tef0 = 0.0 gtfe = 1.0
+ thcs = 0.0 ahc = 0.1 fthc = 0.0 rci0 = 150 vlim = 0.5 vces = 0.1
+ vpt = 0.0 tr = 0.0 cbepar = 0.0 cbcp = 0.0 alqf = 0.0 alit = 0.0
+ flnqs = 0 kf = 0.0 af = 2.0 cfbe = -1 latb = 0.0 latl = 0.0 vgb = 1.17
+ alt0 = 0.0 kt0 = 0.0 zetaci = 0.0 alvs = 0.0 alces = 0.0 zetarbi = 0.0
+ zetarbx = 0.0 zetarcx = 0.0 zetare = 0.0 zetacx = 1.0 vge = 1.17
+ vgc = 1.17 vgs = 1.17 flvg = -1.02377E-4 f2vg = 4.3215E-4
+ zetact = 3.0 zetabet = 3.5 flsh = 0 rth = 0.0 cth = 0.0 flcomp = 0.0
+ tnom = 27.0 dt = 0.0
ends hicumL2_model_default

```

Fig. 8.27: Model card example specifying all model parameters (as default values here).  
This model card can be used to run HICUM/L2 verilog-A code in spectre.

Many HICUM parameters have been chosen as simple factors, that are related to physically meaningful basic parameters like a capacitance, charge or transit time. This choice significantly reduces changes (and the probability of errors) in the parameter list if the basic parameters are changed for, e.g., statistical simulation, because the factors often assume very similar values even for different process technologies. For production-type parameter library releases, it is recommended to have self-heating and non-quasi-static effects turned off by their respective flags, since these effects are not relevant for many design tasks, especially not for first

phase design and feasibility studies. Including these effects will unnecessarily increase the simulation time for *all* users during the entire design cycle. These effects can become important in the last phase for tuning the performance of certain types of circuits or for design verification before tape-out. It is suggested to make these flags available in the design system as options for circuit designers.

In most simulators, a factor  $M$  allows to scale model parameters for identical transistors connected in parallel. Caution is required if this factor is applied to  $R_{su}$ ,  $C_{su}$ ,  $R_{th}$ , and  $C_{th}$ , since it neglects interaction between parallel devices.

As reference temperature, 27°C has been chosen to remain compatible with other simulators and models but this can of course be changed by the user. The value for “ $\infty$ ” may be dependent on the simulator. Since the parameter list below is a snapshot in time capturing the status of a living model that is constantly being improved and adapted to the latest process development, model users are strongly advised to consult the circuit simulator description of the actual model version they are simulating with.

An example for a “model card” is shown in Fig. 8.27, with default values inserted for the parameters, which should be the same in all circuit simulators. With these values, all but the absolutely necessary functions defining a bipolar transistor are turned off so that the user only needs to specify the parameters for those effects that are desired to be taken into account.

name	description	range	unit	factor
	transfer current			
<b>is</b> [c10]	Saturation current (GICCR constant) (c10=is*qp0)	[0:1] ([0:1])	A (AC)	M (M <sup>2</sup> )
<b>qp0</b>	Zero-bias hole charge	(0:1]	C	M
<b>ich</b>	High-current correction for 2D and 3D effects	[0:inf)	A	M
<b>hfe</b>	Emitter minority charge weighting factor in HBTs	[0:inf]	-	



<b>name</b>	<b>description</b>	<b>range</b>	<b>unit</b>	<b>factor</b>
<b>hfc</b>	Collector minority charge weighting factor in HBTs	[0:inf]	-	
<b>hjei</b>	BE depletion charge weighting factor in HBTs	[0:100]	-	
<b>hjei</b>	BC depletion charge weighting factor in HBTs	[0:100]	-	
<b>mcf</b>	Non-ideality factor (used for III-V HBTs)	(0:10]	-	
	Base-emitter current components			
<b>ibeis</b>	Internal BE saturation current	[0:1]	A	M
<b>mbei</b>	Internal BE current ideality factor	(0:10]	-	
<b>ireis</b>	Internal BE recombination saturation current	[0:1]	A	M
<b>mrei</b>	Internal BE recombination current ideality factor	(0:10]	-	
<b>ibeps</b>	Peripheral BE saturation current	[0:1]	A	M
<b>mbep</b>	Peripheral BE current ideality factor	(0:10]	-	
<b>ireps</b>	Peripheral BE recombination saturation current	[0:1]	A	M
<b>mrep</b>	Peripheral BE recombination current ideality factor	(0:10]	-	

<b>name</b>	<b>description</b>	<b>range</b>	<b>unit</b>	<b>factor</b>
<b>tbhrec</b>	Base current recombination time constant at the BC barrier for high forward injection (default is v2.1 compatible)	(0:inf]	s	
	Base-collector currents			
<b>ibcis</b>	Internal BC saturation current	[0:1]	A	M
<b>mbci</b>	Internal BC current ideality factor	(0:10]	-	
<b>ibcx</b>	External BC saturation current	[0:1]	A	M
<b>mbcx</b>	External BC current ideality factor	(0:10]	-	
	Base-collector avalanche current			
<b>favl</b>	Avalanche current factor	[0:inf)	1/V	
<b>qavl</b>	Exponent factor for avalanche current	[0:inf)	C	M
	Vertical non-quasi-static effects			
<b>alqf</b>	Factor for additional delay time of minority charge	[0:1]	-	
<b>alit</b>	Factor for additional delay time of transfer current	[0:1]	-	
<b>flnqs</b>	Flag for turning on (1) or off (0) vertical NQS effects	[0/1]	-	
	Base-emitter tunneling current			
<b>ibets</b>	BE tunneling saturation current	[0:1]	A	M

<b>name</b>	<b>description</b>	<b>range</b>	<b>unit</b>	<b>factor</b>
<b>abet</b>	Exponent factor for tunneling current	[0:inf)	-	
<b>tunode</b>	Specifies the base node connection of the tunneling current source	[0/1]	-	
	Depletion charges and capacitances			
<b>cjei0</b>	Internal BE zero-bias depletion capacitance	[0:inf)	F	M
<b>vdei</b>	Internal B-E built-in potential	(0:10]	V	
<b>zei</b>	Internal BE grading coefficient	(0:1]	-	
<b>ajei</b>	Ratio of maximum to zero-bias value of internal BE capacitance	(0:inf]	-	
<b>cjep0</b>	Peripheral BE zero-bias depletion capacitance	[0:inf)	F	M
<b>vdep</b>	Peripheral BE built-in potential	(0:10]	V	
<b>zep</b>	Peripheral BE grading coefficient	(0:1]	-	
<b>ajep</b>	Ratio of maximum to zero-bias value of peripheral BE capacitance	[0:inf)	-	
<b>cjci0</b>	Internal BC zero-bias depletion capacitance	[0:inf)	F	M
<b>vdci</b>	Internal BC built-in potential	(0:10]	V	
<b>zci</b>	Internal BC grading coefficient	(0:1]	-	

<b>name</b>	<b>description</b>	<b>range</b>	<b>unit</b>	<b>factor</b>
<b>vptci</b>	Internal BC punch-through voltage	(0:100]	V	
<b>cjcx0</b>	External BC zero-bias depletion capacitance	[0:inf)	F	M
<b>vdex</b>	External BC built-in potential	(0:10]	V	
<b>zcx</b>	External BC grading coefficient	(0:1]	-	
<b>vptex</b>	External BC punch-through voltage	(0:100]	V	
<b>fbc</b>	Partitioning factor for external BC capacitance	[0:1]	-	
<b>cjs0</b>	CS zero-bias depletion capacitance	[0:inf)	F	M
<b>vds</b>	CS built-in potential	(0:10]	V	
<b>zs</b>	CS grading coefficient	(0:1]	-	
<b>vpts</b>	CS punch-through voltage	(0:100]	V	
	Minority charge storage effects			
<b>t0</b>	Low-current forward transit time at $V_{BC}=0V$	[0:inf)	s	
<b>dt0h</b>	Time constant for base and BC space charge layer width modulation	[0:inf)	s	
<b>tbvl</b>	Time constant for modelling carrier jam at low $V_{CE}$	[0:inf)	s	
<b>tef0</b>	Neutral emitter storage time	[0:inf)	s	

name	description	range	unit	factor
<b>gtfe</b>	Exponent factor for current dependence of neutral emitter storage time	(0:10]	-	
<b>thcs</b>	Saturation time constant at high current densities	[0:inf)	s	
<b>ahc</b>	Smoothing factor for current dependent of base and collector transit time	(0:10]	-	
<b>ftbc</b>	Partitioning factor for base and collector portion	[0:1]	-	
<b>rci0</b>	Internal collector resistance at low electric field	(0:inf)	$\Omega$	1/M
<b>vlim</b>	Voltage separating ohmic and saturation velocity regime	(0:10]	V	
<b>vces</b>	Internal CE saturation voltage	[0:1]	V	
<b>vpt</b>	Collector punch-through voltage	(0:inf]	V	
<b>bick</b>	Velocity-field dependence exponent in $I_{CK}$	[1:2]	s	
<b>vcbar</b>	Barrier height in the BC region	[0:1]	V	
<b>betacbar</b>	Exponent factor for current dependence of BC barrier height	[1:10]	-	
<b>tbfs</b>	Saturation time constant for BC barrier dependent base charge and transit time	[0:inf)	s	
<b>tr</b>	Storage time for inverse operation	[0:inf)	s	

name	description	range	unit	factor
	Lateral geometry scaling (at high current densities)			
<b>latb</b>	Scaling factor for collector minority charge in direction of emitter width $b_E$	[0:inf)	-	
<b>latl</b>	Scaling factor for collector minority charge in direction of emitter length $l_E$	[0:inf)	-	
	Series resistances			
<b>rbi0</b>	Zero-bias internal base resistance	[0:inf)	$\Omega$	1/M
<b>rbx</b>	External base series resistance	[0:inf)	$\Omega$	1/M
<b>fgeo</b>	Factor for geometry dependence of emitter current crowding ( $r_{Bi}$ )	[0:inf]	-	
<b>fdqr0</b>	Correction factor for modulation by BE and BC Space charge layer	[-0.5:100]	-	
<b>ferbi</b>	Ratio of HF shunt to total internal capacitance (lateral NQS effect)	[0:1]	-	
<b>fqi</b>	Ratio of internal to total minority charge	[0:1]	-	
<b>re</b>	Emitter series resistance	[0:inf)	$\Omega$	1/M
<b>rcx</b>	External collector series resistance	[0:inf)	$\Omega$	1/M
	Substrate transistor			

<b>name</b>	<b>description</b>	<b>range</b>	<b>unit</b>	<b>factor</b>
<b>itss</b>	Saturation current of substrate transistor transfer current	[0:1]	A	M
<b>msf</b>	Forward ideality factor of substrate transfer current	(0:10]	-	
<b>iscs</b>	Saturation current of CS diode	[0:1]	A	M
<b>msc</b>	Ideality factor of CS diode	(0:10]	-	
<b>tsf</b>	Transit time (forward operation)	[0:inf)	s	
	Intra-device substrate coupling			
<b>rsu</b>	Substrate series resistance	[0:inf)	$\Omega$	1/M
<b>csu</b>	Shunt capacitance (caused by substrate permittivity)	[0:inf)	F	M
	Parasitic isolation capacitances			
<b>cbepar</b>	Total parasitic BE capacitance (spacer and metal component)	[0:inf)	F	M
<b>fbepar</b>	partitioning factor of parasitic BE capacitance	[0:1]	-	
<b>cbepar</b>	Total parasitic BC capacitance (trench and metal component)	[0:inf)	F	M
<b>fbepar</b>	Partitioning factor of parasitic BC cap (default is v2.1 compatible)	[0:1]	-	
	Noise			
<b>kf</b>	Flicker noise coefficient (no unit only for AF = 2)	[0:inf)	-	$M^{1-AF}$
<b>af</b>	Flicker noise exponent factor	(0:10]	-	

name	description	range	unit	factor
	Temperature dependence			
<b>vgb</b>	Bandgap-voltage extrapolated to 0K	(0:10]	V	
<b>f1vg</b>	Coefficient $K_1$ in T dependent bandgap equation		V/K	
<b>f2vg</b>	Coefficient $K_2$ in T dependent bandgap equation		V/K	
<b>zetact</b>	Exponent coefficient in transfer current temperature dependence	[-10:10]	-	
<b>kqp0</b>	Coefficient for modeling T-dependence of zero-bias hole charge	[0:1]	-	
<b>vge</b>	Effective emitter bandgap voltage $V_{gEeff}$	(0:10]	V	
<b>zetabet</b>	Exponent coefficient in BE junction current temperature dependence	[-10:10]	-	
<b>vgc</b>	Eff. collector bandgap voltage $V_{gCeff}$	(0:10]	V	
<b>vgs</b>	Eff. substrate bandgap voltage $V_{gSeff}$	(0:10]	V	
<b>alt0</b>	First-order relative temperature coefficient of parameter $t_0$	[-10:10]	1/K	
<b>kt0</b>	Second-order relative temperature coefficient of parameter $t_0$	[-10:10]	1/K <sup>2</sup>	
<b>zetaci</b>	Temperature exponent for $r_{Ci0}$	[-10:10]	-	



name	description	range	unit	factor
<b>alvs</b>	Relative temperature coefficient of saturation drift velocity	[-10:10]	1/K	
<b>alces</b>	Relative temperature coefficient of $V_{CEs}$	[-10:10]	1/K	
<b>zetarbi</b>	Temperature exponent of internal base resistance	[-10:10]	-	
<b>zetarbx</b>	Temperature exponent of external base resistance	[-10:10]	-	
<b>zetarcx</b>	Temperature exponent of external collector resistance	[-10:10]	-	
<b>zetare</b>	Temperature exponent of emitter resistance	[-10:10]	-	
<b>zetacx</b>	Temperature exponent of the mobility in substrate transistor transit time	[-10:10]	-	
<b>alfav</b>	Relative temperature coefficient for $F_{AVL}$	[-10:10]	1/K	
<b>alqav</b>	Relative temperature coefficient for $Q_{AVL}$	[-10:10]	1/K	
	Self-Heating			
<b>rth</b>	Thermal resistance	[0:inf)	K/W	1/M
<b>cth</b>	Thermal capacitance	[0:inf)	Ws/ K	M
<b>altht</b>	Thermal resistance T coefficient	[0:inf)	-	
<b>althp</b>	Thermal resistance power coefficient	[0:inf)	1/W	

name	description	range	unit	factor
<b>flsh</b>	flag for turning on (1,2) or off (0) self-heating effects	[0/1/2]	-	M
	circuit simulator specific parameters			
<b>tnom</b>	Reference temperature $T_0$ at which model parameters are specified		°C	°C
<b>dt</b>	Temperature change w.r.t. chip (substrate) for this particular transistor (assuming substrate temperature to be a global variable)		°C	°C
<b>version</b>	model version identifier	-	-	-

Table 8.2: List of HICUM/L2 model parameters and flags. Note, that all flags, the bandgap voltage parameters (**f1vg**, **f2vg**), and the simulator parameters in the last block are not counted as model parameters. The parameters TNOM and DT are available in most simulators and are also mostly named the same. The “model version identifier” enables version control in simulators with different HICUM generations.

## 8.8 References

- [1] [http://www.iese.tu-dresden.de/iese/eb/hic\\_new/hic\\_start.html](http://www.iese.tu-dresden.de/iese/eb/hic_new/hic_start.html)
- [2] A. Koldehoff, M. Schröter, and H.-M. Rein, “A compact bipolar transistor model for very high-frequency applications with special regard to narrow stripes and high current densities”, *Solid-State Electron.*, Vol. 36, pp. 1035-1048, 1993.
- [3] M. Schröter and T.-Y. Lee., “A physics-based minority charge and transit time model for bipolar transistors”, *IEEE Trans. Electron Dev.*, Vol. 46, pp. 288-300, 1999.
- [4] P.B. Weil und L.P. McNamee, “Simulation of Excess Phase in Bipolar Transistors”, *IEEE Trans. Circ. Syst.*, Vol. CAS-25, 1978, S. 114-116.

- [5] C. McAndrew et al., “BJT Modeling with VBIC, Basics and V1.3 Updates”, and private communication, 2007.
- [6] G. Coram, private communication, 2007.
- [7] M. Schröter and H. Tran, “Modeling of charge and collector field in Si-based bipolar transistors”, *Proc. WCM, International NanoTech Meeting, Boston (MA)*, pp. 102-107, 2004.
- [8] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener and A. Koldehoff, “Physics- and process-based bipolar transistor modeling for integrated circuit design”, *IEEE Journal of Solid-State Circuits*, Vol. 34, pp. 1136-1149, 1999.
- [9] M. Rudolph and F. Bonani, “Low-frequency noise in nonlinear systems”, *IEEE Microwave Magazine*, p. 84-92, Feb. 2009.
- [10] J Paasschens, R. Havens, L. Tiemeijer, “Modelling the Correlation in the High-Frequency Noise of (Heterojunction) Bipolar Transistors using Charge-Partitioning”, *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 221-224, 2003.
- [11] P. Sakalas, A. Chakravorty, J. Herricht, M. Schröter, “Compact Modeling of High Frequency Correlated Noise in HBTs”, *Proc. Bipolar Circuits and Technology Meeting (BCTM), Maastricht (Belgium)*, pp. 279-282, 2006.
- [12] A. Van der Ziel and G. Bosman, “Noise temperature of common emitter microwave transistors”, *IEEE Trans. Electron Dev.*, Vol. 31, pp. 1280-1283, 1984.

This page is intentionally left blank

## **Chapter 9**

### **Parameter Determination for HICUM/L2**

The potential of a compact model can never be fully exploited without proper parameter determination. This includes the overall strategy (or methodology) with particular general and model-specific methods for obtaining the model parameter values as well as the execution of those methods (usually called extraction). Depending on the methodology, such as physics-based scalable extraction or direct parameter fitting, special test structures in addition to regular transistors may be used. As a consequence, compact model parameter extraction is quite a wide area that fills an entire book on its own. Therefore, this chapter can only provide a rough overview on the basic ideas and principles rather than a detailed description of each possible method.

## 9.1 Introduction

It is understood that HICUM/L2 is capable of addressing the modeling issues related to the design of modern integrated circuits. To cover the large variety of applications the number of required transistor configurations may span a very wide range. For example, in case of high-speed circuits, it is common that the emitter window width  $b_{E0}$  is equal to the minimum manufacturable value  $b_{E0min}$ , and the optimum emitter length  $l_{E0}$  is much larger than  $b_{E0}$ , while in case of driver circuits and power amplifiers the optimum  $b_{E0}$  can be larger than  $b_{E0min}$  [1,2]. On the other hand, low-power circuits often require transistors whose emitter size approaches the lower manufacturable limit, i.e.  $l_{E0} \rightarrow b_{E0} = b_{E0min}$ . Furthermore, in radio-frequency (RF) circuits, performance optimization and impedance matching require the capability to select the proper transistor size to achieve the necessary trade-offs [3]. For this, a suitable procedure for calculating the optimum emitter size and transistor configuration for the particular application is needed [1,4,5]. The variety of applications requires libraries with typically more than 100 different configurations. Unfortunately, the desired large number of transistor configurations is often difficult to support because the commonly used bipolar transistor model parameter extraction methods rely on fitting individual transistors to measurement data. Therefore, conventional parameter determination becomes rather time consuming and expensive for a large number of configurations, limiting the size of a library and, hence, circuit optimization and performance.

Often realistic model parameters are needed already in the early phase of process development, in which suitable data or even wafer material do not exist. In addition, for economic reasons, process development should be aligned with application needs, often requiring a quick evaluation of the impact of process changes on circuit performance. Just from these examples, it is obvious that the conventional method of placing every (anticipated) library transistor on a test chip, waiting for processing a “near nominal” wafer, and then fitting the model parameters of each transistor is not suited to serve the above needs.

Besides having an accurate (semi-)physical compact model like HICUM/L2, the above-mentioned problems can only be overcome by provid-

ing modeling engineers and circuit designers with a “tool” for generating process-based sets of model parameters for arbitrary transistor configurations. For this, sufficiently accurate analytical relations have to be established for model parameters as a function of transistor dimensions as well as length- or area-specific process and electrical data, which can be measured easily using adequate test and transistor structures. The Transistor Dimensioning and Calculation (TRADICA) computer program ([5-8]) has been developed and constantly improved since the early eighties to address these issues. Functionality and capabilities of TRADICA are explained in sec. 11.1.

For HICUM/L2, geometry and process data in conjunction with special test structures are used to obtain as many as possible input data independently and, hence, to make the model parameters as physics-based as possible. In addition to bias, frequency and temperature, the transistor geometry is considered as an independent variable (vector) for parameter extraction. The use of geometry information helps avoiding ambiguous values compared to just fitting the terminal characteristics of a single device. In this chapter the basic extracting sequence is described that can be used for generating geometry scalable HICUM/L2 parameters. For additional information on the extraction procedure and commercially available tools see, e.g., [9-13].

Absolute parameters are converted into a geometry and layout independent form to obtain the so-called specific electrical data, which are then used in TRADICA to generate model parameters for arbitrary transistor configurations. For a given process, obtaining compact model parameters of a large variety of transistor configurations using the recommended process-based scalable approach (PBSA), involves the following major steps:

- wafer selection according to certain criteria (see next section);
- deriving the relevant transistor dimensions from design rules;
- measurement of the relevant characteristics of a certain set of test structures (including transistors) over bias, geometry, temperature and frequency;
- extraction of (geometry) specific model parameters;
- generation of the model parameters for desired transistor configurations (either as library or directly during the circuit design phase).

The various aspects related to the above steps are briefly discussed in this chapter. An overview on the recommended sequence of parameter extraction flow is given in a formal way in section 9.5 to provide the reader with basic information, such as measurement and data requirements as well as fundamental procedures employed. Another purpose of this overview is also to serve as a guide line for implementing a parameter extraction methodology. The simulator DEVICE [14] is used to obtain the device simulation examples that helped to visualize the important steps.

## 9.2 Wafer selection

When it comes to model parameter extraction there are many criteria as to which wafers should be selected or rejected. Those criteria that proved to be useful for the PBSA are briefly discussed here. In modern bipolar technologies, often at least two “types” of transistors are available: a high-performance (HP) transistor and a high-voltage (HV) transistor, that are defined by the same process flow but with just one additional mask for the selectively implanted collector. Using information obtained for both types can improve the efficiency of the parameter extraction.

The PBSA relaxes the requirements for the wafer selection, since it allows to shift the (specific) parameters later on to their desired nominal values (cf. sec. 11.2.4). However, the electrical performance of the transistors (and other parameters) should not be too far off from the target specifications. Also, it is important to evaluate the process regarding geometry scalability, since non-standard scaling (cf. sec. 5.5) consumes a larger effort (and more time) for parameter extraction. Therefore, before a wafer is accepted for parameter extraction purposes, the following tests are recommended to be performed and evaluated:

- Measurement of a tetrode structure (for each transistor “type” and nominal  $b_{E0}$ ) [15,16] at zero bias, yielding roughly the internal base sheet resistance  $r_{SBi0}$ ;
- Measurement of a large area BC diode (i.e. without SIC) at zero bias and beyond punch through, yielding epi doping  $N_C$  and thickness  $w_C$ ;
- Measurement of the bias dependent S-parameters for a typical transistor (each type) at a single CE voltage and frequency, yielding the transit frequency  $f_T$ .



The first two measurements can be performed on PCM structures, which effectively indicate the range of process variations for device performance. The third one is more time consuming and can be done after the first two have been evaluated. In general, though, all of the above mentioned structures and data are usually already available during process development and evaluation.

It is recommended then to obtain a wafer map that shows the uniformity of the electrical parameters  $r_{S\text{Bi}0}$ ,  $w_C$ ,  $N_C$ , and  $f_T$ . The correlation of the latter to the first three should be checked as well as, of course, the absolute values regarding their deviation with respect to the target values. Based on the wafer map, an appropriate die for parameter extraction is selected from the center of the area that shows the highest uniformity. As a consequence, the chance of deviations in electrical characteristics between different transistors of the selected die is minimized.

In addition to the electrical tests, it is highly recommended to obtain pictures of the cross-section and top view (SEM and TEM pictures) of the most important transistor configurations for both extractions and applications. These pictures are usually available during process development and not only visualize the *actual* transistor structure but also serve for verifying the transistor dimensions assumed or calculated from design rules. The information from the SEM/TEM pictures can avoid non-physical results and geometry scaling problems. The latter cannot be explained otherwise just from electrical measurements.

### 9.3 Relevant transistor dimensions

Definitions of the relevant transistor dimensions used for calculating the area and perimeter length specific model parameters are given in Fig. 9.1 for the example of a SiGe HBT. The described parameter extraction procedure can also be applied to BJTs.

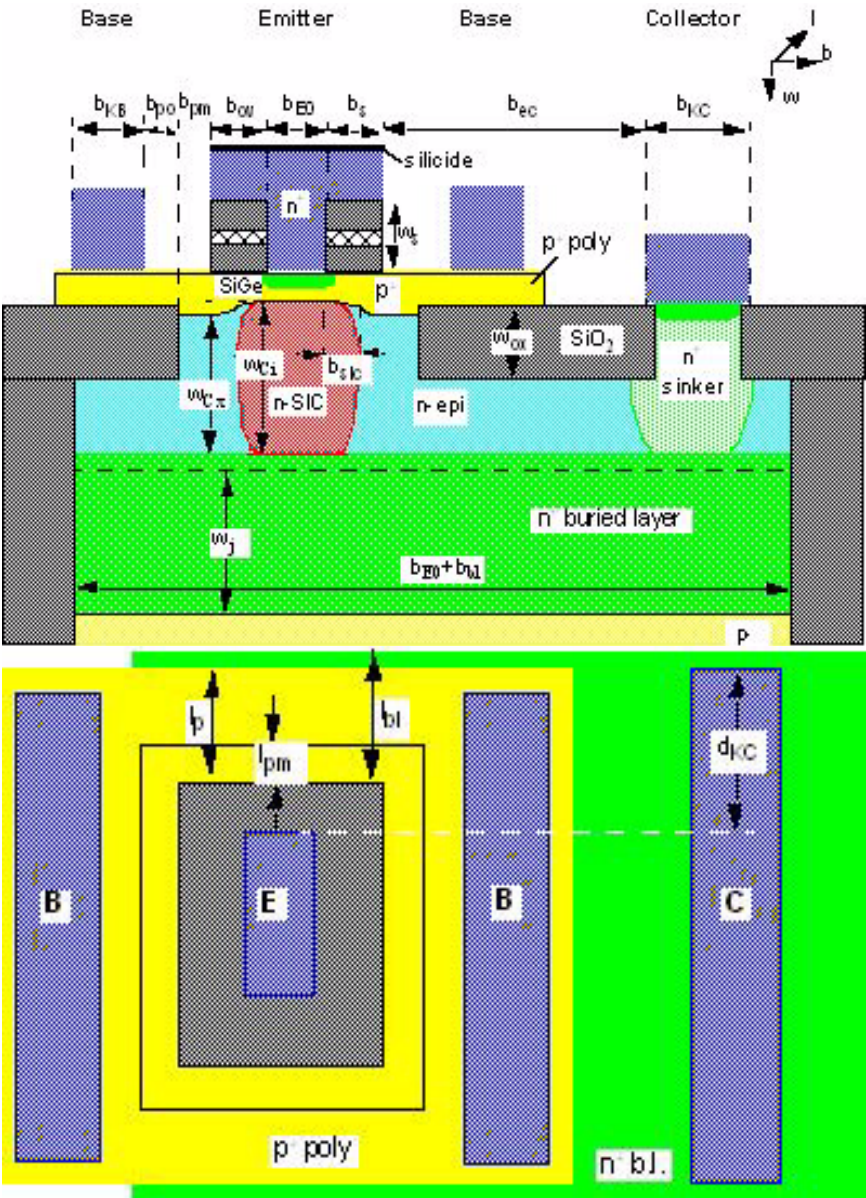


Fig. 9.1: Schematic cross-section and layout of an epitaxial base SiGe bipolar transistor with shallow and deep trench isolation.

## 9.4 Measurements

Parameter extraction for the PBSA relies on minimum requirements regarding measurement effort and equipment. While the set-up and detailed bias conditions (test plans) for a particular parameter determination procedure are given in the respective sections below, at this point a brief overview shall be provided on the basic and most important measurement methods, which are being employed for obtaining the data required for a variety of procedures

The measurements are generally taken on transistors and special test structures. Examples of the most important test structures suitable for PBSA are given in [15, 16, 17]. A description of details of measurements and data acquisition is beyond the scope of this book. Readers interested in further information are referred to e.g. [18].

### 9.4.1 IV measurements and data

DC measurements taken with a parameter analyzer on test structures and transistors, are called IV measurements, resulting in the associated IV data. Examples are:

- a forced current in a 3- or 4-terminal test structure (e.g. tetrode, Kelvin structures) for determining a sheet resistance and the corresponding measured voltage between the contacts or
- the collector and base current of a transistor as function of the applied voltage.

However, IV data are also included in the DC bias taken during S-parameter measurements (see below), while *IV measurements* always means DC operation and the associated set-up as defined above.

### 9.4.2 CV measurements and data

Capacitors with a sufficiently large value can be measured with LCR or CV meters, thus the name CV measurements and CV data. This equipment usually operates at frequencies between 0.1...10 MHz. It is, therefore, only suited for measuring large size capacitors that are laid out as special test structures like in process control monitors.

The designation CV data includes capacitance versus voltage data from both CV measurements, as defined above, and from (cold) S-parameter measurements. For advanced SiGe HBT processes, large area BE and BC diodes may not yield the same area-specific value as narrow structures, if the vertical doping profile changes with emitter size. Hence, it is recommended to determine transistor relevant CV data from S-parameters on structures within the size range that is allowed for circuit design.

### 9.4.3 S-parameter measurements and data

Small-signal measurements at high frequencies are performed with a vector network analyzer (VNA), resulting in S-parameter data. Operating frequencies for obtaining accurate data are usually above 300 MHz, with an upper limit usually in the 26...220 GHz range, dependent on the respective equipment and investigated process technology. For parameter extraction purposes, two types of S-parameter measurements are often distinguished: (a) “cold” measurements during which the transistor is operated at reverse bias up to very low forward bias with negligible carrier injection and current across the junctions; (b) “hot” measurements during which the transistor is operated under usual forward-bias conditions with significant carrier injection and non-negligible current across junctions. The corresponding results are usually designated as, e.g. cold S-parameters or hot S-parameters data.

Although the operating frequencies of high-performance transistors in some of the present Si/SiGe-based processes reach several 100GHz, the corresponding parameters for modeling the high-frequency transistor behavior can be determined at frequencies well below 100 GHz under so-called quasi-static conditions. The lower frequency limit for parameter extraction related S-parameter measurements is approximately given by the 3-dB corner frequency of the common-emitter (CE) small-signal current gain. At current gains between about 50 and 400, this corresponds to a range of 2... 10 GHz. One of the most important quasi-static small-signal figures of merit (FoMs) for extracting model parameters is the transit frequency  $f_T$  of the transistor. It has been shown in [19] that  $f_T$  can be obtained from a single (spot) frequency measurement of the small-signal CE current gain.

A typical set of data from a single-frequency bias sweep measurement includes the terminal voltages and currents as well as the frequency and the four S-parameters in either magnitude and phase or real and imaginary part representation. The S-parameters can then be converted into Y-parameters at the same bias points (e.g. [20]). Next, the Y-parameters need to be de-embedded in order to eliminate the influence of parasitic elements and to obtain the Y-parameters of the device under test only [21-24]. From the de-embedded Y-parameters as a function of bias, the HF behavior related HICUM/L2 model parameters can be extracted.

Generally, the bias dependence of the small-signal characteristics and transistor parameters (such as the transconductance) is of great interest for model parameter extraction. In many circuit design applications, the dependence of small-signal transistor parameters on the collector current  $I_C$  at a given voltage  $V_{CE}$  or  $V_{CB}$  is of major importance. Unfortunately, it is found very often, that S-parameters as a function of frequency are taken for bias points, that are defined by the terminal voltages  $V_{BE}$  and  $V_{CE}$ , while the terminal currents are not or cannot be monitored. The latter is often caused by, e.g., limitations of the data acquisition software or the equipment. The terminal currents are then obtained with a separate DC measurement in an attempt to establish the relation between S-parameters and bias currents. The great danger in this procedure is that self-heating usually causes the devices to heat up differently during the S-parameter and the DC measurement, leading to an incorrect relationship of the above mentioned variables. The consistent way of taking data is to at least monitor the current or, better, to define the bias point by the collector current. The most suitable method is to take iso-thermal data using pulsed-bias measurements.

#### 9.4.4 Measurement conditions

Table 9.1 contains an overview on the most important measurements to be performed and the associated bias conditions. The values are examples and are given for Si/SiGe-based processes. Only a minimum number of measurements is specified but more data is always useful.

Measurement type and bias conditions	Data	Result
<ul style="list-style-type: none"> <li>internal base pinch resistance data from at least 3 tetrodes with different <math>b_{E0}</math>;               <ul style="list-style-type: none"> <li><math>V_{BE}=[-0.5,0.5]\text{V}</math> @ <math>V_{CE}=0</math>, <math>\Delta V_{BE}=0.1\text{V}</math>, <math>\Delta V_{BB}=0.01\text{V}</math></li> </ul> </li> <li>sheet and contact resistances of external base region, buried layer, collector (sinker, contact)               <ul style="list-style-type: none"> <li><math>\Delta V=0.01\dots0.1\text{V}</math>, depending on resistance value</li> </ul> </li> </ul>	$V_{BE} I_{B1} I_{B2}$  $\Delta V I$ (or $r_S$ , $r_{con} \dots$ )	base and collector series resistance components
<ul style="list-style-type: none"> <li>C-V on large area transistor               <ul style="list-style-type: none"> <li><math>V_{BE}=[-0.5,0.5]\text{V}</math>, <math>V_{BC}=V_{SC}=0</math>, <math>\Delta V_{BE}=0.1\text{V}</math></li> <li><math>V_{BC}=[-BV_{CEO},0.5]\text{V}</math>, <math>V_{BE}=V_{SC}=0</math>, <math>\Delta V_{BC}=0.1\text{V}</math></li> <li><math>V_{SC}=[-BV_{CEO},0.5]\text{V}</math>, <math>V_{BE}=V_{BC}=0</math>, <math>\Delta V_{SC}=0.1\text{V}</math></li> </ul> </li> <li>C-V on large-area BC diode (only for structures without selectively implanted collector)               <ul style="list-style-type: none"> <li><math>V_{BC}=[-BV_{CBO},0.5]\text{V}</math>, <math>V_{BE}=V_{SC}=0</math>, <math>\Delta V_{BC}=0.1\text{V}</math></li> </ul> </li> </ul>	$V_{BE} C_{jE}$ $V_{BC} C_{jC}$ $V_{SC} C_{jS}$  $V_{BC}$ $C_{jC}(\text{epi})$	depletion and isolation capacitance components   breakdown voltage, collector doping and width
<ul style="list-style-type: none"> <li>cold S-parameters as a function of bias on <math>\geq 3</math> transistors with different <math>b_E</math> (<math>\ll l_{E0}</math>)               <ul style="list-style-type: none"> <li><math>V_{BE}=[-0.5, 0.5]\text{V}</math>, <math>V_{BC}=0</math>, <math>\Delta V_{BE}=0.1\text{V}</math></li> <li><math>V_{BC}=[-BV_{CEO}, 0.5]\text{V}</math>, <math>V_{BE}=0</math>, <math>\Delta V_{BC}=0.1\text{V}</math></li> </ul> </li> </ul>	$V_{BE} \underline{S}$ $V_{BC} \underline{S}$	depletion and isolation capacitance components

Measurement type and bias conditions	Data	Result
<ul style="list-style-type: none"> <li>S-parameters as a function of bias on at least 3 transistors with different <math>b_{E0}</math> (<math>\ll I_{E0}</math>):             <ul style="list-style-type: none"> <li><math>I_C/A_E=[0.01, J_{Cp}] \text{mA}/\mu\text{m}^2</math> (depends on process) for at least 3 <math>V_{CE}</math>, e.g. <math>V_{CE}/V=0.5, 1.5, BV_{CEO}</math></li> </ul> </li> </ul>	$V_{BE} \ V_{CE} \ I_C$ $I_B \ \underline{S}$	$f_T, \tau_f$ ; certain forward I-V parameters; verification
<ul style="list-style-type: none"> <li>DC output characteristics (isothermal measurement at <math>T_0</math> for reference transistor only):             <ul style="list-style-type: none"> <li><math>V_{CE}=[0V, V_{CE,max}]</math> @ <math>I_B=\text{const}, V_{BE}=\text{const}</math> (<math>V_{CE,max} &lt; BV_{CEO}</math> (high <math>I_C/A_{E0}</math>))</li> <li><math>I_C/A_{E0}=[0.01, J_{Cp}] \text{mA}/\mu\text{m}^2</math> (depends on process)</li> </ul> </li> </ul>	$V_{CE} \ I_C \ I_B$ $V_{BE}$	Avalanche current, certain $I_C$ parameters; verification
<ul style="list-style-type: none"> <li>DC reverse characteristic (isothermal measurement at <math>T_0</math> for reference transistor only):             <ul style="list-style-type: none"> <li><math>V_{BC}=[0.4, 0.7]V</math> @ <math>V_{BE}=0V</math></li> </ul> </li> </ul>	$V_{BC} \ I_B \ [I_C]$	BC diode current
<ul style="list-style-type: none"> <li>Temperature dependence: e.g. <math>T=[-40, 75, 125]^\circ\text{C}</math> <ul style="list-style-type: none"> <li>repeat above measurements</li> <li>shift <math>V_{BE}</math> bias according to temperature, assuming a TC of about <math>-1.5\text{mV/K}</math></li> </ul> </li> </ul>	$T, \dots$	TCs

Table 9.1: Typical measurement conditions for bipolar transistors.  $J_{Gp} = 1.5J_C(f_{Tpeak})$  at  $V_{CE,max}$ . The thermal limit for the devices may be obtained from the measurement of temperature dependence.  $A_{E0}$  is the emitter window area,  $T_0$  is the reference temperature.

### 9.4.5 Sequence of measurements

In a production environment, measurement effort and time have to be minimized. This is done on one hand by using standard and established equipment set-ups and on the other hand by maximizing the equipment utilization in a given time frame.

Data acquisition for model parameter extraction starts at the reference temperature  $T_0$  with simple and fast IV measurements of all special test structures used for determining sheet and contact resistances. Next, CV measurements are performed, followed by single-frequency cold and hot S-parameter measurements. The same device is probed for all bias conditions before moving the probes to the next device. These measurements provide already sufficient information for extracting more than 80% of the (specific) model parameters.

In the meantime, i.e. during parameter extraction, data acquisition can continue with temperature dependent measurements, which are quite time consuming. By the time the latter measurements are completed, parameter extraction has provided an overview on the process and its actual performance, so that those bias ranges and points can be determined, which are of interest for frequency sweeps. Frequency sweep measurements are usually required and often taken only at the reference temperature. Based on the temperature and frequency dependent data, the remaining model parameters (except those for low-frequency  $1/f$  noise) can be extracted, and model verification can already start. Finally, special measurements, such as those for noise and distortion, are performed for further, application specific, model validation.

## 9.5 Extraction flow

Parameter extraction often turns out to be the bottleneck for introducing new methods or new compact models. Therefore, it is felt that a good understanding of the extraction methodology is necessary to also obtain a feeling of the capabilities and potential of the practical deployment of a compact model.

The parameter-extraction procedure depends to a certain extent on the chosen compact model. Table 9.2 gives an overview on the HICUM/L2 re-



lated parameter-extraction sequence. During most steps, data from devices with different sizes are treated simultaneously. Optimization and error compensation are avoided as much as possible by using linearly independent measured data. The program TRADICA allows then to generate HICUM/L2 model parameters for any transistor structure from a single set of extracted geometry specific parameters.

Step	Flow of Parameter Extraction
1	Junction capacitances (including bottom/periphery separation) and device inherent parasitic (spacer/oxide) capacitances from S-parameter and LCR measurements <ul style="list-style-type: none"> <li>• BC depletion capacitance components and related punch-through voltages</li> <li>• BE depletion capacitance components</li> <li>• CS depletion capacitance components</li> </ul>
2	<ul style="list-style-type: none"> <li>• internal base sheet resistance</li> <li>• sheet resistances and specific contact resistance of external base region</li> <li>• buried layer sheet and collector sinker/contact resistance</li> <li>• specific emitter resistance (e.g. combined with thermal resistance)</li> </ul>
3	Geometry dependence of saturation current densities (collector and base currents) including bottom/periphery separation and related model parameters from bias dependent DC IV characteristics at low injection
4	<ul style="list-style-type: none"> <li>• temperature coefficients (IV, CV, transit time), needed for corrections</li> <li>• thermal resistance</li> </ul>
5	Geometry dependence of low to medium current $f_T$ <ul style="list-style-type: none"> <li>• current independent transit time and minority charge from (spot frequency) S-parameter measurements</li> <li>• parameters related to the critical current <math>I_{CK}</math></li> </ul>

Step	Flow of Parameter Extraction
6	Geometry dependence of high-current $f_T$ : <ul style="list-style-type: none"> <li>transit time and minority charge (including lateral scaling) from bias dependent (spot frequency) S-parameter measurements</li> <li>forward bias base-emitter depletion capacitance</li> </ul>
7	avalanche breakdown model parameters
8	low frequency 1/f noise parameters
9	parasitic substrate transistor parameters
10.	High frequency effects: <ul style="list-style-type: none"> <li>BC capacitance splitting (<math>f_{BCpar}</math>)</li> <li>substrate network (<math>R_{su}</math>, <math>C_{su}</math>)</li> <li>quasi-static effects from <math>y_{21}</math> and <math>y_{11}</math> (<math>\alpha_{IT}</math>, <math>\alpha_{Qf}</math>)</li> <li>high-frequency noise parameters</li> </ul>
11.	process tolerances for independent process parameters
12.	first verification and possible fine tuning based on y-parameters vs. frequency, bias, geometry, and temperature
13.	final verification: high frequency noise, distortion, benchmark circuits

Table 9.2: Sequence for process-based (geometry) scalable model parameter extraction for HICUM/L2, assuming that prior to step 1 detailed information on design rules and device geometries is available.

## 9.6 Step-by-step extraction procedure

The goal of this section is to give a “formal” overview on the sequence for extracting geometry scalable HICUM/L2 parameters. To keep this overview efficient, for each step only a brief description of the applied extraction procedure is provided, while for a detailed description (including equations) and the background of the respective procedure the reader is referred to references and previous chapter(s) of this book. During develop-

ing HICUM/L2, it has always been kept in mind that model equations should enable a parameter extraction procedure in which as many as possible steps can be performed linearly independently or with only a weak interdependence, particularly for determining parameters describing first-order effects. For commercially available implementations see [11, 12], and a comprehensive set of test structures for industrial purposes was presented in [17].

It is recommended to extract first the parameters of the HV transistors and then the parameters of the HP transistors since the collector is playing a dominating role in HV transistors affecting the overall characteristics. If HV transistors are not offered, those ones required for extraction can usually be realized easily and should be included on a test chip.

A couple of assumptions are being made in order to apply the procedures in practice:

- A suitable wafer with the appropriate test structures and transistors has been selected that has passed the recommended acceptance check described in section 9.2.
- All measurements that are required for a particular extraction step are available and have been properly de-embedded. It is generally preferable to convert S-parameters to Y-parameters (to be done during de-embedding in any way) and use the latter data for parameter extraction.
- The design rules and dimensions of all test structures and transistors are known and have been verified, so that all necessary geometry calculations can be performed.
- The process is geometry scalable, i.e. the profile under the emitter does not depend significantly on the emitter width.

The information about each extraction step is given in a modular way in the following sub-sections. The meaning of the key words on the right-hand-side and the terminology used shall be briefly explained below.

- “Data” characterizes the type of data required for the particular step. Acronyms such as CV, IV or cold measurements have already been defined in section 9.4.
- “Model parameters” specify those ones that are needed for the present step and had to be extracted in an earlier step. They do not include dimensions, which are specified under “required geometry data”.

- “Procedure” refers to the main extraction method and required steps.
- Under “Extracted specific parameters” those parameters are listed that are geometry independent. These parameters are used in TRADICA to generate geometry scalable libraries for model parameters. Of course, certain HICUM/L2 parameters, such as ratios, are geometry independent in the first place and do not need to be scaled with geometry. However, since it is useful to determine a full set of (geometry) specific parameters first and keep the data in one place prior to subsequent parameter (library) generation, all extracted parameters are listed here.
- “Related HICUM/L2 parameters” are those that eventually are written into a library (as model card) and are generated for a particular transistor configuration, employing a program like TRADICA.

Figure 9.2 shows a sketch of the forward-bias collector and base current characteristics with important parameters having an influence on the DC characteristics.

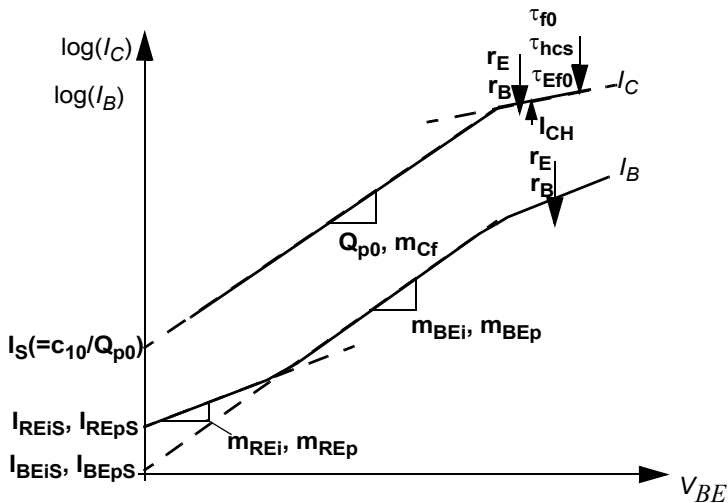


Fig. 9.2: HICUM/L2 parameters related to forward DC collector and base current characteristics. Arrows indicate the impact of an increase of the respective parameter.

A graphical presentation of HICUM/L2 depletion capacitance related model parameter extraction is roughly outlined in Fig. 9.3. Parameters re-

lated to the transit time are usually determined from de-embedded S-parameter data via  $1/2\pi f_T$  versus  $1/I_C$  plot as shown in Fig. 9.4.

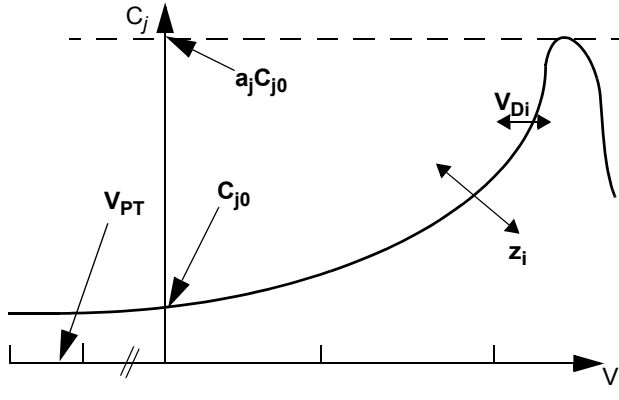


Fig. 9.3: HICUM/L2 parameters related to the depletion capacitance from CV characteristics. Reverse bias region scale is modified.

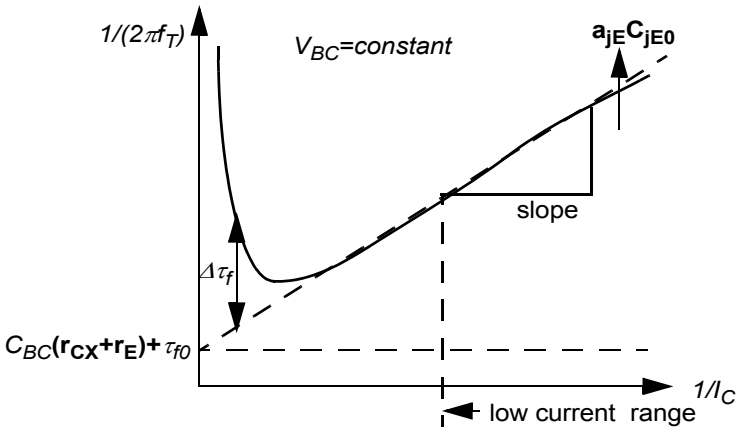


Fig. 9.4: HICUM/L2 parameters related to the transit time from the reciprocal of the cut-off frequency vs. reciprocal of collector current (usually determined from S-parameter data).

To visualize the extraction methods outlined below, 2-D device simulation results of SiGe HBTs (under isothermal condition at 300K) with different geometries were used, in which the emitter window length  $l_{E0}$  was much larger than the emitter width  $b_{E0}$ , which varies from 0.2  $\mu\text{m}$  to

0.6  $\mu\text{m}$ . Figure 9.5 shows the device structure with  $b_{E0} = 0.3 \mu\text{m}$  and its doping profile is similar to one given in Fig. 2.3b.

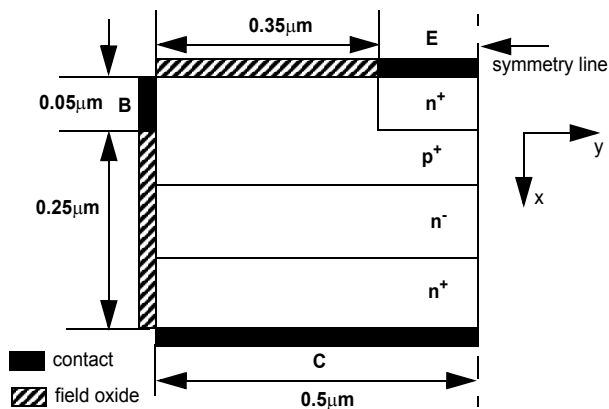


Fig. 9.5: Schematic cross-section for 2D SiGe-HBT structure with  $b_{E0} = 0.3 \mu\text{m}$ . For symmetry reason only half of the structure is shown.

In the following, linearly independent steps within the extraction methodology are indicated by the entry “none” under “Required model parameters”.

### 9.6.1 BC depletion and isolation capacitance

The capacitance can be determined from direct CV measurements or cold-S parameters.

Data	<ul style="list-style-type: none"><li>• cold Y-parameters of different transistor configurations</li><li>• CV data of large area HV structure (optional)</li><li>• CV data of large area HP structure</li></ul>
Required model parameters	none (optional: $C_{BCpar}$ e.g. for each transistor configuration calculated from TRADICA, based on measurements or on device simulation)
Required geometry data	<ul style="list-style-type: none"><li>• area <math>A_{BC}</math> and perimeter <math>P_{BC}</math> of BC junction (HV transistor)</li><li>• area <math>A_{SIC}</math> of SIC region (HP transistor)</li></ul>

Procedure:	<ul style="list-style-type: none"> <li>• HV data: separation into bias dependent bottom and perimeter specific components of the depletion capacitance via different geometries; also, determination of specific isolation capacitance possible.</li> <li>• Extraction of parameters for modeling the bias dependence of above depletion capacitances.</li> <li>• HP CV data: extraction of SIC related parameters.</li> </ul>
Extracted (specific) parameters	<ul style="list-style-type: none"> <li>• HV data: <math>C_{jC0,A}</math>, <math>V_{DCa}</math>, <math>z_{Ca}</math>, <math>V_{PTCa}</math>; <math>C_{jC0,P}</math>, <math>V_{DCp}</math>, <math>z_{Cp}</math>, <math>V_{PTCp}</math>, <math>[C_{BCpar,P}]</math></li> <li>• HP data: <math>C_{jCi0,A}</math>, <math>V_{DCi}</math>, <math>z_{Ci}</math>, <math>V_{PTCi}</math></li> </ul>
Related HICUM parameters	$C_{jCx0}$ , $V_{DCx}$ , $z_{Cx}$ , $V_{PTCx}$ , $C_{BCpar}$ , $f_{BCpar}$ ; $C_{jCi0}$ , $V_{DCi}$ , $z_{Ci}$ , $V_{PTCi}$ (cf. Fig. 9.6)

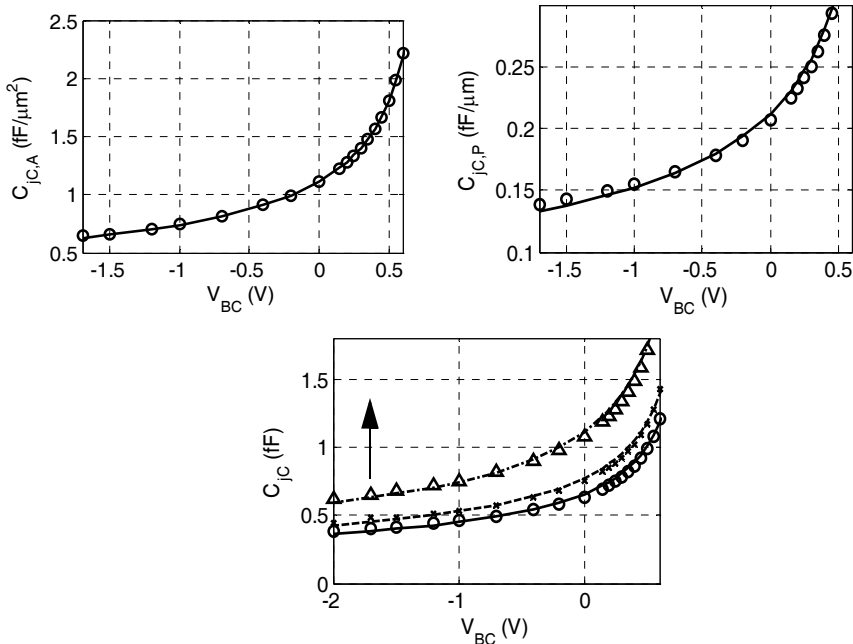


Fig. 9.6: BC depletion capacitances (symbols for device simulation and lines for HICUM/L2): bottom components (top left), peripheral components (top right), and total capacitance (bottom) with  $b_{E0}/\mu\text{m} = 0.2, 0.3, \text{ and } 0.6$ .

### 9.6.2 BE depletion and isolation capacitance

Both the internal and peripheral BE capacitance are important for the investigated structures. Bottom and peripheral components are decoupled and modelled separately to obtain the excellent agreement in Fig. 9.7.

Data, test structures	<ul style="list-style-type: none"> <li>• cold and hot Y-parameters of different transistor configurations</li> <li>• CV data of large area structure (optional)</li> </ul>
Model parameters	none (optional: $C_{BEpar}$ e.g. for each transistor configuration calculated from TRADICA, based on measurements or on device simulation)
Required geometry data	area $A_{E0}$ and perimeter $P_{E0}$ of emitter window
Procedure:	<ul style="list-style-type: none"> <li>• Determination of <math>C_{BE}</math> from <math>1/(2\pi f_T)</math> at a forward bias point.</li> <li>• Separation into bias dependent bottom and perimeter specific components of the depletion capacitance via different geometries; also, determination of specific isolation capacitance possible.</li> <li>• Extraction of parameters for modeling the bias dependence of above depletion capacitances [10,25]</li> </ul>
Extracted (specific) parameters	$C_{jEi0,A}, V_{DEi}, z_{Ei}, a_{jEi}, C_{jEp0,P}, V_{DEp}, z_{Ep}, a_{jEp}, [C_{BEpar,P}]$
Related HICUM parameters	$C_{jEi0}, V_{DEi}, z_{Ei}, a_{jEi}, C_{jEp0}, V_{DEp}, z_{Ep}, a_{jEp}, C_{BEpar}$ (cf. Fig. 9.7)



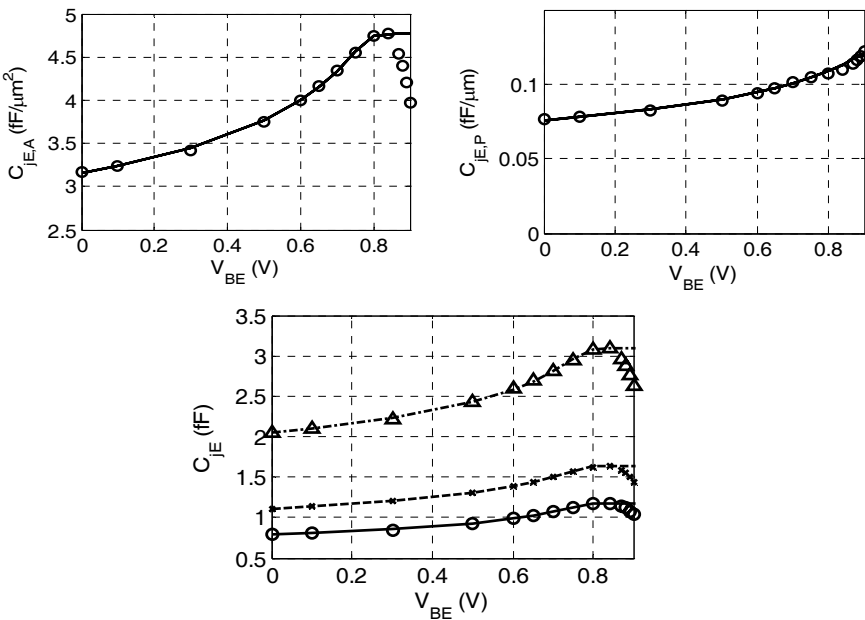


Fig. 9.7: BE depletion capacitances (symbols for device simulation and lines for HICUM/L2): bottom components (top left), peripheral components (top right), and total capacitance (bottom) with  $b_{E0}/\mu m = 0.2, 0.3$ , and  $0.6$ .

### 9.6.3 CS depletion capacitance

Since no substrate is present in the example structures, no separate plots are provided for CS capacitance modeling.

Data and test structures	<ul style="list-style-type: none"><li>• cold Y-parameters of different transistor configurations</li><li>• CV data of large area structure (optional, but recommended)</li></ul>
Model parameters	none
Required geometry data	area $A_{CS}$ and perimeter $P_{CS}$ of CS junction

Procedure:	<ul style="list-style-type: none"> <li>• Separation into bias dependent bottom and perimeter specific components of the depletion capacitance via different geometries.</li> <li>• Extraction of parameters for modeling the bias dependence of above depletion capacitances.</li> </ul>
Extracted (specific) parameters	$C_{jS0,A}, V_{DSa}, z_{Sa}, V_{PTSa}, C_{jS0,P}, V_{DSp}, z_{Sp}, V_{PTSp}$
Related HICUM parameters	$C_{jS0}, V_{DS}, z_S, V_{PTS}$

#### 9.6.4 Internal base (sheet) resistance

The internal zero-bias and bias-dependent hole charge are crucial for modelling the bias-dependent internal base-sheet resistance. Figure 9.8 shows the accuracy of the model equation plotted against the device simulation results. The extracted value for  $Q_{p0} = 19.8$  fC agrees quite well with the actual value (18.4 fC) directly obtained from device simulation. Further improvement is possible by setting  $V_{CE} = 0$  instead of  $V_{BC} = 0$  [15].

Data	<p>IV data on transistor tetrodes with different emitter widths:</p> <ul style="list-style-type: none"> <li>• sweep of <math>V_{BE} = V_{BC}</math> (<math>V_{CE} = 0</math>)</li> <li>• sweep of <math>V_{BE}</math> @ <math>V_{BC} = 0</math></li> </ul>
Model parameters	$C_{jEi0}, V_{DEi}, z_{Ei}, C_{jCi0}, V_{DCi}, z_{Ci}$ (or numerical integration of associated depletion charges)
Required geometry data	width $b_{E0}$ and length $l_{E0}$ of emitter windows
Procedure:	<ul style="list-style-type: none"> <li>• Determine internal base sheet resistance <math>r_{SBi}(V_{BE}, V_{BC})</math> from corrected data [15,16].</li> <li>• Extraction of parameters for modeling the bias dependence of <math>r_{SBi}</math>.</li> <li>• Determination of the link and total external resistance (used for next step)</li> </ul>

Extracted (specific) parameters	$r_{SBi0}, Q_{p0,A}, f_{dQr0}, r_{Ss}$
Related HICUM parameters	$R_{Bi0}, Q_{p0}, f_{dQr0}$ . The parameter $f_{geo}$ can be directly calculated from the transistor configuration

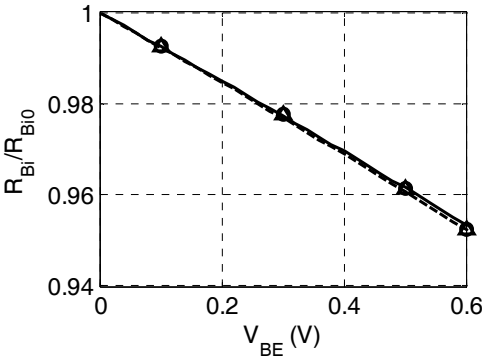


Fig. 9.8: Bias dependence of normalized internal base resistance at  $V_{BC} = 0V$  with  $b_{E0}/\mu m = 0.2$ , and  $0.6$ . Symbols represent device simulation, lines for HICUM/L2.

9.6.5 Components of external base resistance

The extraction and corresponding scalability was studied in, e.g., [26].

Data	<ul style="list-style-type: none"><li>• IV data of various resistance test structures</li><li>• link and total external resistance from transistor test-rodes</li></ul>
Model parameters	None
Required geometry data	dimensions of the relevant regions of the test structures
Procedure:	<ul style="list-style-type: none"><li>• Determine resistance(s) from IV data and perform current spreading correction (depending on resistance type).</li><li>• Extract specific contact and sheet resistance from each structure</li></ul>

Extracted (specific) parameters	$\rho_{KB}, r_{Spo}, r_{Spm}, r_{Ssil}$
Related HICUM parameter	$R_{Bx}$

**9.6.6 Emitter resistance**

Emitter resistance is crucial for a number of characteristics as shown in Fig. 9.2. Therefore, its correct determination is very important. The improved open-collector method [27] has been found to give reasonable and reliable results. However, the current flow through the emitter window deviates from the one in usual transistor operation. The Extraction method based on the transconductance [26] is found to be another attractive option. Its results are plotted in Fig. 9.9.

Data	open collector IV data from transistor structures with different emitter size
Model parameters	None
Required geometry data	total emitter window width $A_{E0}$ of each transistor
Procedure:	<ul style="list-style-type: none"><li>• for each transistor: fit modified open-collector model equation to measured data</li><li>• extract specific contact resistance from <math>r_E(1/A_{E0})</math></li></ul>
Extracted (specific) parameter	$\rho_{KE}$
Related HICUM parameter	$R_E$ (cf. Fig. 9.9)

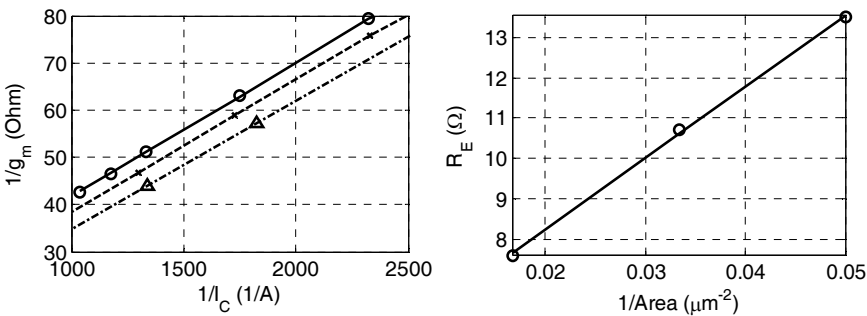


Fig. 9.9: Emitter resistance,  $r_E$ , extracted using the method mentioned in [26] at  $V_{BC} = 0V$  with  $b_{E0}/\mu m = 0.2, 0.3, 0.6$  (left), and its geometry scaling (right). Symbols represent extracted values, lines linear regression fit results.

9.6.7 Components of external collector resistance

The PBSA, which is based on a special test-structure, is outlined in the following table. For a single transistor structure other methods such as, e.g., the optimization of output characteristics in hard saturation region [26] are required.

Data	IV data of special resistance test structure
Model parameters	None
Required geometry data	dimensions of the relevant regions of the test structure
Procedure:	<ul style="list-style-type: none"><li>• Determine resistance from IV data and perform current spreading correction (if required).</li><li>• Extract buried layer sheet and specific contact (incl. sinker) resistance</li></ul>
Extracted (specific) parameters	$r_{Sbl}, \rho_{KC}$
Related HICUM parameter	$R_{Cx}$

### 9.6.8 Collector current at low bias

A scalable approach is outlined below and corresponding results are plotted in Figs. 9.10 and 9.11. The extraction of low injection collector current parameters for a single transistor structure can be found in literature.

Data	IV data from transistors with different emitter size
Model parameters	most processes: $C_{jEi0}$ , $V_{DEi}$ , $z_{Ei}$ , $a_{jEi}$ optional: $Q_{p0}$ ; [ $C_{jCi0}$ , $V_{DCi}$ , $z_{Ci}$ , if $V_{CE} = \text{const}$ )]
Required geometry data	emitter window area of the transistors
Procedure:	<ul style="list-style-type: none"> <li>• Separation into bias dependent bottom and perimeter specific current components via different geometries [28]</li> <li>• Extraction of parameters related to bias dependence from least-squares fit of <math>\log(I_C/A_E)</math> vs. <math>V_{BE}</math> @ <math>V_{BC}=0</math> (procedure is partially dependent on process)</li> </ul>
Extracted (specific) parameters	$\gamma_C$ , $I_{S,A}$ , [ $Q_{p0,A}$ , $m_{Cf}$ ]
Related HICUM parameters	$I_S$ (or $c_{I0}$ ), [ $Q_{p0}$ , $m_{Cf}$ ] (cf. Figs. 9.10 and 9.11)

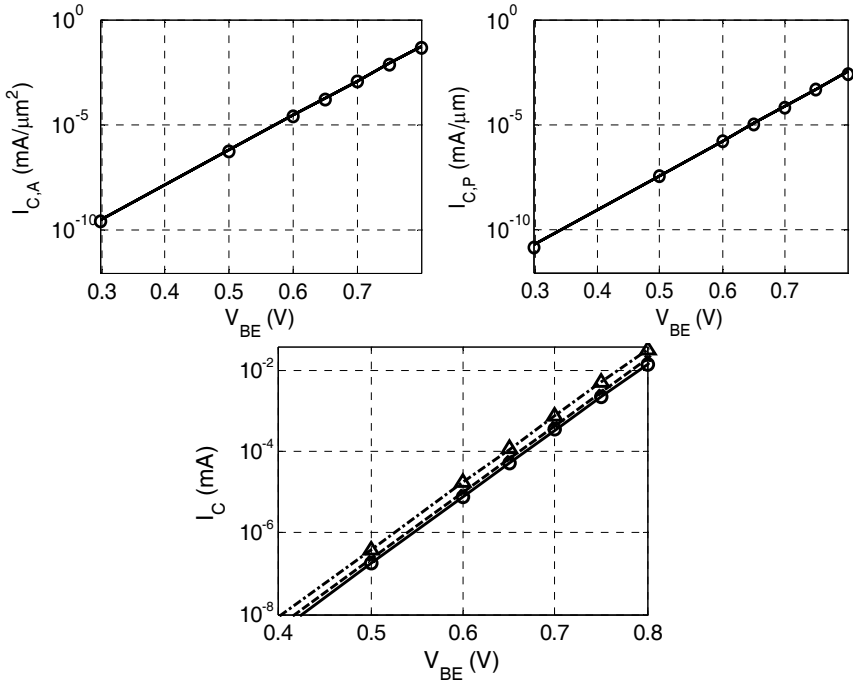


Fig. 9.10: Bias dependence of collector current at low-injection at  $V_{BC} = 0V$  (symbols for device simulation and lines for HICUM/L2): bottom components (top left), peripheral components (top right), and total current (bottom) with  $b_{E0}/\mu m = 0.2, 0.3$ , and  $0.6$ .

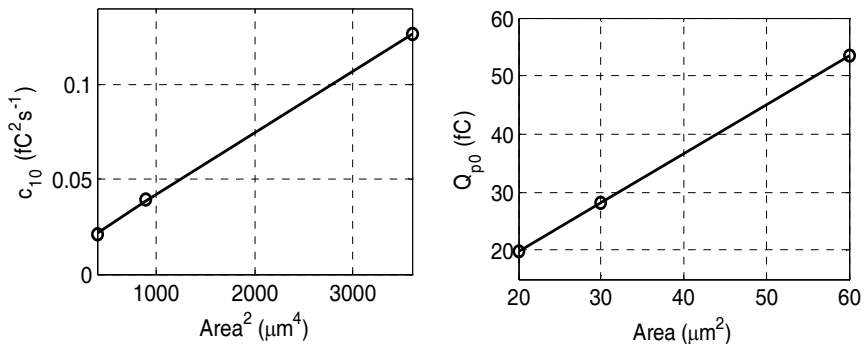


Fig. 9.11: Geometry dependence of HICUM/L2 parameters  $c_{10}$  and  $Q_{p0}$  for collector current at low-injection. Symbols represent extracted values, lines linear regression fit results.

### 9.6.9 Current across BE junction at low bias

The recombination and backinjection related parameters are obtained via area and perimeter specific parameters following the steps given in the following table. Corresponding results are plotted in Fig. 9.12.

Data	IV data from transistors with different emitter size
Model parameters	none
Required geometry data	emitter window area of the transistors
Procedure:	<ul style="list-style-type: none"> <li>• Separation into bias dependent bottom and perimeter specific current components via different geometries.</li> <li>• Extraction of saturation current (density) and non-ideality coefficient for each component, e.g., from least-squares fit of <math>\log(I)</math> vs. <math>V_{BE}</math> @ <math>V_{BC}=0</math></li> </ul>
Extracted (specific) parameters	$\gamma_B$ ; $I_{BEiS,A}$ , $m_{BEi}$ , $I_{BEpS,P}$ , $m_{BEp}$ ; $I_{REiS,A}$ , $m_{REi}$ , $I_{REpS,P}$ , $m_{REp}$
Related HICUM parameters	$I_{BEiS}$ , $m_{BEi}$ , $I_{BEpS}$ , $m_{BEp}$ ; $I_{REiS}$ , $m_{REi}$ , $I_{REpS}$ , $m_{REp}$ (cf. Fig. 9.12)



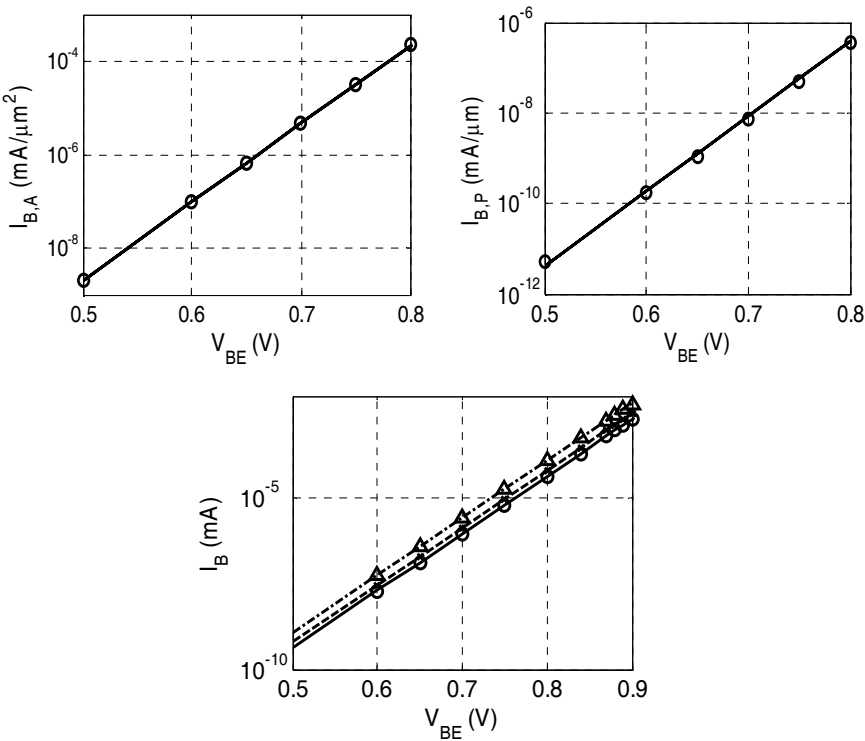


Fig. 9.12: Bias dependence of base-emitter current at low-injection at  $V_{BC} = 0\text{V}$  (symbols for device simulation and lines for HICUM/L2): bottom components (top left), peripheral components (top right), and total current (bottom) with  $b_{E0}/\mu\text{m} = 0.2, 0.3$ , and  $0.6$ .

### 9.6.10 Current across the BC junction (at low bias)

For long transistor structures, the area of the BC junction hardly changes with emitter width so that the BC diode current depends on weakly on geometry as shown in Fig. 9.13.

Data	IV data from transistors with different size
Model parameters	none

Required geometry data	collector-base junction area $A_{BC}$ and emitter window area $A_{E0}$ of the transistors
Procedure:	<ul style="list-style-type: none"><li>• Separation into bias dependent bottom and perimeter specific current components via different geometries.</li><li>• Extraction of saturation current (density) and non-ideality coefficient for each component, e.g., from least-squares fit of <math>\log(I)</math> vs. <math>V_{BC}</math> @ <math>V_{BE}=0</math></li></ul>
Extracted (specific) parameters	$I_{BCxS,A}, m_{BCx}; I_{BCiS,A}, m_{BCi}$
Related HICUM parameters	$I_{BCxS}, m_{BCx}; I_{BCiS}, m_{BCi}$ (cf. Fig. 9.13)

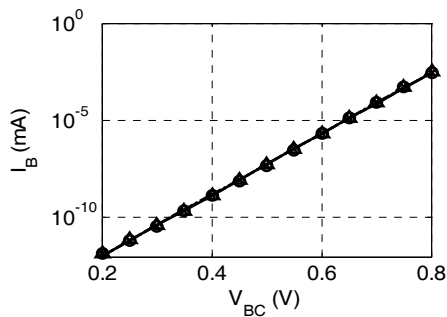


Fig. 9.13: Bias dependence of base-collector current in “inverse” operation at  $V_{BE} = 0V$  with  $b_{E0}/\mu m = 0.2, 0.3$ , and  $0.6$ . Symbols represent device simulation, lines for HICUM/L2.

9.6.11 Thermal resistance

Preferably with a prior extraction of parasitic resistances, e.g.,  $R_E$ , thermal resistance can be extracted following the table below.

Data	IV data of transistors used for extraction
------	--

Model parameters	$R_E, R_{Bx}, R_{Bi}$ (dependent on method)
Required geometry data	emitter window area $A_{E0}$ of the transistors
Procedure:	<ul style="list-style-type: none"> <li>Extraction of <math>R_{th}</math> of each transistor from <math>I_B</math> as a function of dissipated power according to [29], but with known <math>R_E</math>.</li> </ul> <p>For alternative methods see e.g. [30, 31]</p>
Extracted (specific) parameter	$r_{th}$
Related HICUM parameter	$R_{th}$

### 9.6.12 Forward transit time

This is one of the most critical steps in HICUM/L2 parameter extraction. It consists of determining first the parameter of the low current transit time, than those of the critical current, and lastly those of the high-current transit time. Determining the critical current is quite crucial, since the accuracy of its extraction ultimately determines the high-frequency model accuracy. In Fig. 9.14 relevant extraction results are plotted.

Data	hot Y-parameters of different transistor configurations
Model parameters	$C_{jCi0}, V_{DCi}, z_{Ci}, V_{PTCi}; C_{jCx0}, V_{DCx}, z_{Cx}, V_{PTCx},$ $C_{BCpar};$ $r_E, r_{Cx}, \gamma_C; [R_{th}]$
Required geometry data	emitter window dimensions $b_{E0}$ and $l_{E0}$ of the transistors

Procedure:	<ul style="list-style-type: none"> <li>• Determine transit frequency <math>f_T</math> from Y-parameters and determine transit time <math>\tau_f</math> from <math>1/(2\pi f_T)</math> vs <math>1/I_C</math>.</li> <li>• Low-current range: <ul style="list-style-type: none"> <li>• From <math>\tau_{f0}(V_{BC})</math> data, extract parameters describing the voltage dependence.</li> <li>• Determine bottom and perimeter related component and associated geometry factor from <math>\tau_{f0}(V_{BC}=0)</math> of transistors with different emitter size.</li> </ul> </li> <li>• Medium current range <ul style="list-style-type: none"> <li>• From <math>I_{CK}(V_{CE}</math> or <math>V_{BC})</math> data, extract parameters describing the current dependence</li> <li>• Extract current spreading angle from <math>I_{CK}(V_{CE}=0.8V</math> or <math>V_{BC}=0V)</math></li> </ul> </li> <li>• High-current region: extract relevant parameters describing the bias dependence [9, 10, 32]</li> </ul>
Extracted (specific) parameters	$\tau_{f0i}, f_{tpi} = \tau_{f0p}/\tau_{f0i}, \Delta\tau_{0h}, \tau_{Bfvl}, r_{Ci0,A}, V_{lim}, V_{PT}, V_{CEs}; a_{hc}, \tau_{hcs}, \tau_{Ej0}, g_{tE}, f_{thc}$
Related HICUM parameters	$\tau_0, \Delta\tau_{0h}, \tau_{Bfvl}, r_{Ci0}, V_{lim}, V_{PT}, V_{CEs}; a_{hc}, \tau_{hcs}, \tau_{Ej0}, g_{tE}, f_{thc}$ (cf. Fig. 9.14)

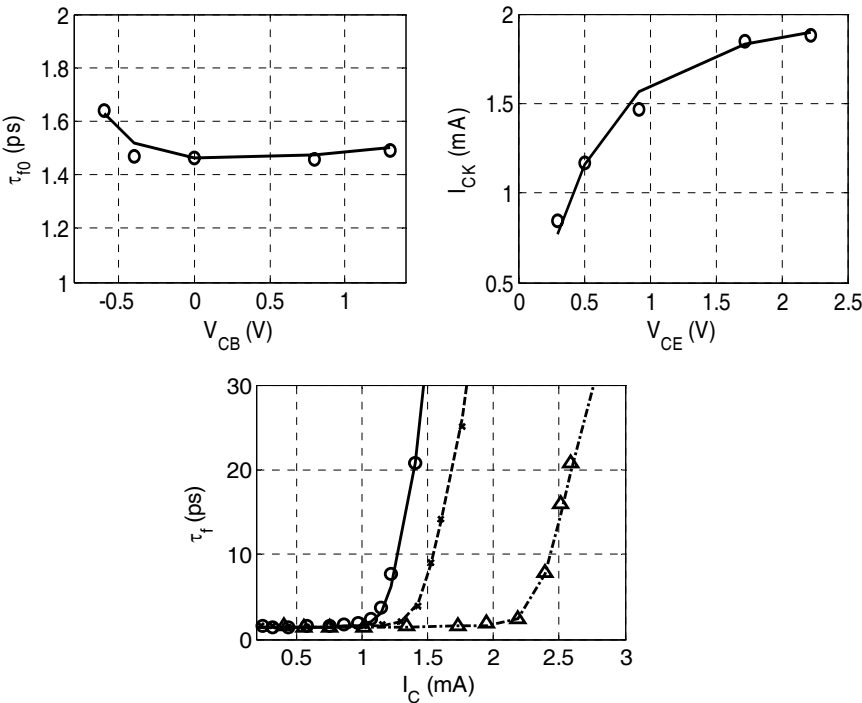


Fig. 9.14: Low current transit time (top left), critical current (top right) for SiGe-HBT with  $b_{E0}/\mu m = 0.3$  (symbols for device simulation and lines for HICUM/L2), and total transit time (bottom) with  $b_{E0}/\mu m = 0.2, 0.3$ , and  $0.6$  at  $V_{BC} = 0V$ .

9.6.13 Collector current at high injection

Getting good agreement for the collector current at high injection using the extracted physics-based model parameters seems to be a fairly straightforward task at this point. However, errors in the required model parameters listed in the following table will lead to deviations of the model characteristics. The model comparison with device simulation is shown in Fig. 9.15.

Data	IV data from transistors with different emitter size
------	--

Model parameters	<ul style="list-style-type: none"><li>• those for <math>I_T</math> and <math>I_{BE}</math> extracted at low injection</li><li>• depletion capacitances and transit time (to calculate <math>Q_{p,T}</math>)</li><li>• <math>r_E, r_{Bx}, r_{Bi}, r_{Cx}</math></li><li>• <math>R_{th}</math></li></ul>
Required geometry data	emitter window area of the transistors
Procedure:	<ul style="list-style-type: none"><li>• Extraction in high current region via non-linear optimization of <math>\log(I_C/A_E)</math> vs. <math>V_{BE}</math> at sufficiently low <math>V_{CE}</math> (to minimize impact of self-heating).</li><li>• Optimization can be performed simultaneously on transistors with different emitter sizes.</li></ul>
Extracted (specific) parameter	$I_{Ch,A}$
Related HICUM parameter	$I_{Ch}$ (cf. Fig. 9.15)

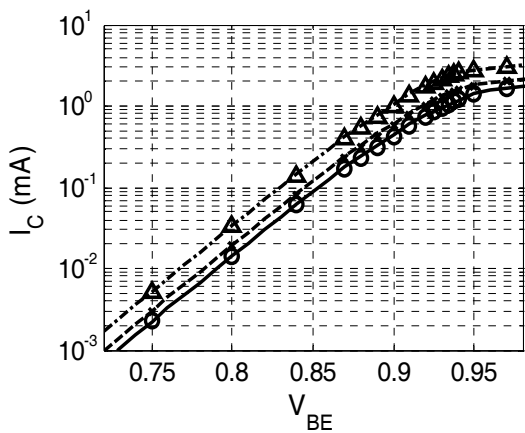


Fig. 9.15: Collector current at high injection for  $b_{E0}/\mu\text{m} = 0.2, 0.3$ , and  $0.6$  at  $V_{BC} = 0\text{V}$ . Symbols for device simulation and lines for HICUM/L2.

### 9.6.14 Base-collector Breakdown

Only weak avalanche “local” breakdown is considered in HICUM/L2. Using the procedure in the following table, scalable values of the relevant parameters are obtained. The model is compared with device simulation data in Fig. 9.16.

Data	$I_B(V_{CB} \text{ or } V_{CE})$ data from transistors with different size
Model parameters	$C_{jCi0}, V_{DCi}, z_{Ci}, V_{PTCi}, \gamma_C$
Required geometry data	emitter window area $A_{E0}$ of the transistors
Procedure:	<ul style="list-style-type: none"> <li>• Determination of avalanche current <math>I_{AVL}(V_{CB})</math> from measured <math>I_B(V_{CB})</math> data at sufficiently low forward bias <math>V_{BE}</math>.</li> <li>• Extraction of parameters describing the bias dependence via non-linear optimization of <math>I_{AVL}(V_{CB})</math> [33].</li> </ul>
Extracted (specific) parameters	$f_{AVL}, q_{AVL,A}$
Related HICUM parameters	$f_{AVL}, q_{AVL}$ (cf. Fig. 9.16)

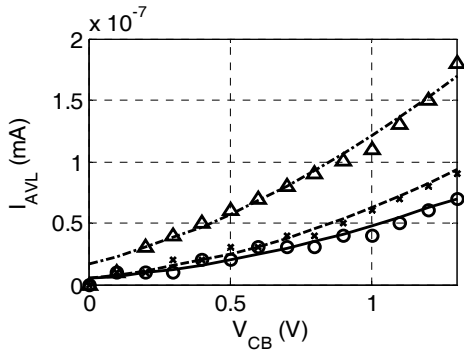


Fig. 9.16: Avalanche current vs. BC voltage for  $b_{E0} = 0.2, 0.3, \text{ and } 0.6 \mu m$ . symbols for device simulation and lines for HICUM/L2.

9.6.15 High-Frequency effects

Under HF conditions NQS effects and proper partitioning of capacitances (and associated charges) across mainly  $R_{Bx}$  becomes critical.

9.6.15.1 Vertical non-quasi-static effects

The delay times are extracted according to the procedure given in the table below. The results are plotted in Fig. 9.17 considering two cases for the model: without NQS effect ( $\mathbf{f}_{nqs} = 0$ ) and with NQS effect enabled ( $\mathbf{f}_{nqs} = 1$ ). It is found that  $\alpha_{Qf}$  does not have any visible effect on the investigated devices, whereas  $\alpha_{iT}$  has significant effect as shown through the frequency dependent phase of  $Y_{21}$ .

Data	hot Y-parameters of a typical transistor configuration
Model parameters	<ul style="list-style-type: none"><li>those for <math>I_T</math> and <math>I_{BE}</math> extracted at low injection</li><li>all capacitances and transit time</li><li><math>R_E, R_{Bx}, R_{Bi}, R_{Cx}</math></li></ul>
Required geometry data	None



Procedure:	For transistor with minimum emitter width: <ul style="list-style-type: none"><li>• Extraction of <math>\alpha_{iT}</math> by fitting the phase of <math>y_{2I}</math> at high frequencies for various bias points [8].</li><li>• Extraction of <math>\alpha_{Qf}</math> by fitting the phase (or <math>\text{Re}\{y_{1I}\}</math>) at high frequencies for various bias points.</li></ul>
Extracted (specific) parameters	$\alpha_{iT}, \alpha_{Qf}$
Related HICUM parameters	$\alpha_{iT}, \alpha_{Qf}, f_{nqs}$ (cf. Fig. 9.17)

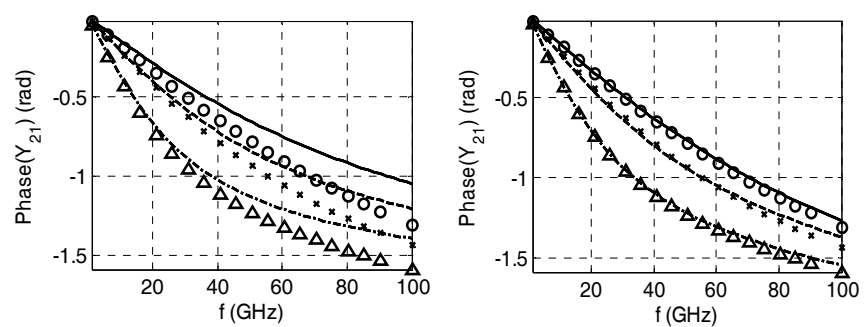


Fig. 9.17: Frequency dependence of phase( $Y_{21}$ ) disabling (left) and enabling (right) vertical NQS effects (through  $\alpha_{iT}$ ) for  $b_{E0}/\mu\text{m} = 0.2, 0.3$ , and  $0.6$  at  $V_{BE} = 0.86\text{ V}$ ,  $V_{BC} = 0\text{ V}$ , and  $f_T = 95\text{ GHz}$ . Symbols correspond to device simulation and lines to HICUM/L2.

9.6.15.2 Partitioning of the external BC capacitance

Presently existing methods, that are based on experimental data, only allow to extract the partitioning factor for a single geometry, but do not offer a generic description for geometry scaling.

Data	None
Model parameters	specific BC capacitances and sheet/contact resistances of the external base

Required geometry data	dimensions of the various regions of the external base
Procedure	Calculation of the partitioning factor by TRADICA for each geometry during parameter (library) generation, or adjustment to power gain or $f_{\max}$
Extracted (specific) parameter	$f_{BCpar}$
Related HICUM parameter	$f_{BCpar}$

#### 9.6.16 Intra-device substrate coupling

Presently existing methods only allow to extract the substrate resistance or substrate network elements for a single geometry (e.g. [34-37]), but do not offer a generic description for geometry scaling.

Data	cold Y-parameters of transistors with relevant configurations or of special test structures.
Model parameters	$R_E$ (is of minor importance though) depends on test structure and method used.
Required geometry data	dimensions of CS junction, deep trench and distance to substrate contact
Procedure:	<ul style="list-style-type: none"> <li>Determine the impedance <math>Z_{su}</math> of the substrate coupling network and extract <math>R_{su}</math> and <math>C_{su}</math> from real and imaginary part of <math>1/Z_{su}</math>.</li> <li>Alternative (also for parameter library generation): calculation from device simulation and TRADICA after experimental calibration.</li> </ul>
Extracted (specific) parameters	$R_{su}$ , $C_{su}$

Related HICUM parameters	$R_{su}, C_{su}$
--------------------------	------------------

### 9.6.17 High-frequency emitter current crowding

This lateral NQS effect is caused by the dynamic charge storage in the internal base region (cf. sec. 8.1.4). Figure 9.18 shows a comparison of the model simulation without considering the effect and with the effect enabled. It is observed that for the device with  $b_{E0} = 0.2\mu\text{m}$  lateral NQS does not have any effect, whereas it is significant for higher dimension devices.

Data	hot Y-parameters of preferably wide transistor configuration(s)
Model parameters	<ul style="list-style-type: none"> <li>• those of <math>i_T</math> and <math>i_{BE}</math></li> <li>• all capacitances (incl. transit time) at the base node</li> <li>• <math>R_E, R_{Bx}, R_{Bi}</math></li> </ul>
Required geometry data	None
Procedure:	for transistor with largest emitter width: <ul style="list-style-type: none"> <li>• <math>f_{CrBi}</math> can be determined from optimizing <math>y_{11}</math> at high frequencies.</li> <li>• Alternative (also for parameter library generation): use theoretical value calculated by TRADICA</li> </ul>
Extracted (specific) parameter	$f_{CrBi}$
Related HICUM parameter	$f_{CrBi}$ (cf. Fig. 9.18)

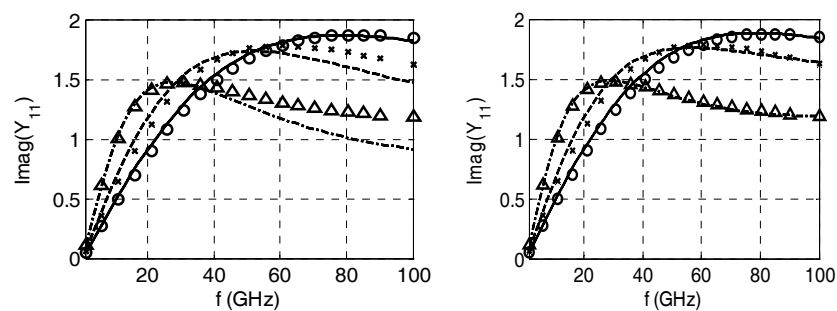


Fig. 9.18: Frequency dependence of  $\text{Imag}(Y_{11})$  disabling (left) and enabling (right) lateral NQS effects (through  $f_{\text{CrBi}}$ ) for  $b_{E0}/\mu\text{m} = 0.2, 0.3$ , and  $0.6$  at  $V_{BE} = 0.86\text{ V}$ ,  $V_{BC} = 0\text{ V}$ , and  $f_T = 95\text{ GHz}$ . Symbols for device simulation and lines for HICUM/L2.

9.6.18 Parasitic substrate transistor elements

For parameter extraction of the substrate transistor related currents a DC test structure with separate substrate contact is preferred. Geometry scaling of the substrate junction related current depends on the process.

9.6.18.1 Transfer current

Extraction procedure is almost similar to the one detailed in section 9.6.9 used to extract parameters related to BE diode currents.

Data	transistor configurations with different substrate size
Model parameters	<ul style="list-style-type: none"><li>• if separate substrate pad is available: none</li><li>• if in HF pads: base current components</li></ul>
Required geometry data	substrate perimeter length and/or area

Procedure:	<ul style="list-style-type: none"> <li>• Separation into bias dependent bottom and/or perimeter specific current components via different geometries.</li> <li>• Extraction of parameters related to bias dependence from least-squares fit of <math>\log(I_{TS})</math> vs. <math>V_{SC}</math>.</li> </ul>
Extracted (specific) parameters	$I_{TS,S,P}$ or $I_{TS,S,A}$ , $m_{sf}$
Related HICUM parameters	$I_{TS,S}$ , $m_{sf}$

### 9.6.18.2 Charge storage time

It is preferable to have the substrate transistor separately available in HF pads to directly measure its S-parameters. Alternatively, a regular transistor in common-emitter configuration can be measured in hard saturation.

Data	hot Y-parameters for a critical transistor configuration at (very) low $V_{CE}$ (optional: different transistor configurations)
Model parameters	transit time, $C_{BE}$ , $C_{BC}$
Required geometry data	none
Procedure:	adjust $\tau_{sf}$ to fit $f_T$ at low $V_{CE}$ .
Extracted (specific) parameter	$\tau_{sf}$
Related HICUM parameter	$\tau_{sf}$

### 9.6.19 Temperature dependence

For extracting the temperature related model parameters, some of the same measurements as for  $T_0$  are repeated for different temperatures  $T$ . An

outline for temperature related model parameter extraction is given here followed by visualizing the modeling of two main characteristics to showcase the thermal modeling capability of HICUM/L2.

### 9.6.19.1 Bandgap or energy gap voltages related parameters

For the coefficients  $f1_{V_g}$  and  $f2_{V_g}$ , that occur in the T-dependent bandgap equation, default values can be used for silicon based processes. Extraction of the bandgap voltages is very important due to bandgap-narrowing effects and material composition. The bandgap parameters are strongly related to the saturation currents. In addition, exponent coefficients need to be extracted to model the temperature dependence of saturation current parameters for the transfer  $I_S$  (or for  $c_{10}$ ) as well as for junction currents ( $I_{xS}$  with  $x = \text{BEi, BEp, REi, REp, BCi, BCx, SC}$ ). Although, all junction current components include a temperature (exponent) coefficient, only two of them are considered as model parameters, since all others can be estimated from the knowledge of the extracted two parameters

Data	$I_C(V_{BE})$ data at $V_{BC}=0$ for different temperature $T$
Model parameters	saturation current components at $T_0$
Required geometry data	none
Procedure:	<ul style="list-style-type: none"> <li>• Determine the saturation current <math>I_S</math> and the non-ideality coefficient at various temperatures.</li> <li>• fitting of the ratio <math>\log[I_S(T)/I_S(T_0)]</math> vs. <math>(T/T_0)</math> with the respective <math>V_{gB}</math> and exponent factor <math>\zeta</math> as parameter.</li> </ul>
Extracted (specific) parameter	$V_{gB}, \zeta_{CT}, \zeta_{BET}, V_{gE}, V_{gC}, V_{gS}$
Related HICUM parameter	$V_{gB}, \zeta_{CT}, \zeta_{BET}, V_{gE}, V_{gC}, V_{gS}$

### 9.6.19.2 Series resistances

Data	IV data at different temperatures $T$
Model parameters	series resistances and/or their components at $T_0$
Required geometry data	None
Procedure:	<ul style="list-style-type: none"> <li>determine the respective resistance <math>R</math> for each <math>T</math></li> <li>fitting of the ratio <math>\log[R(T)/R(T_0)]</math> vs. <math>(T/T_0)</math> with the respective exponent factor <math>\zeta</math> as parameter.</li> </ul>
Extracted (specific) parameters	$\zeta_{Ci}$ , $\zeta_{rBi}$ , $\zeta_{rBx}$ , $\zeta_{rCx}$ , $\zeta_{rE}$
Related HICUM parameters	$\zeta_{Ci}$ , $\zeta_{rBi}$ , $\zeta_{rBx}$ , $\zeta_{rCx}$ , $\zeta_{rE}$

### 9.6.19.3 Transit time at low current densities

Accurate determination of the first and second order relative TCs is essential to obtain good agreement at different temperatures for both  $\tau_f$  and  $f_T$ .

Data	hot Y-parameters vs bias (at $V_{BC}=0$ ) for different temperatures $T$
Model parameters	none
Required geometry data	none
Procedure:	<ul style="list-style-type: none"> <li>Determine the low-current transit time <math>\tau_{f0}</math> for each <math>T</math></li> <li>Extract <math>\alpha_{t0}</math> and <math>k_{t0}</math> from <math>\tau_{f0}(T)</math> via non-linear optimization</li> </ul>
Extracted (specific) parameters	$\alpha_{t0}$ , $k_{t0}$

Related HICUM parameters	$\alpha_{t0}, k_{t0}$
--------------------------	-----------------------

#### 9.6.19.4 Collector-emitter saturation voltage

The relative TC of the internal CE saturation voltage influences the thermal behavior of  $I_{CK}$  and, accordingly, the high-current characteristics.

Data	hot Y-parameters vs bias (at $V_{BC}=0$ ) for different temperatures $T$
Model parameters	none
Required geometry data	none
Procedure	<ul style="list-style-type: none"> <li>• Determine the transit time <math>\tau_f</math> and the critical current <math>I_{CK}</math> for each <math>T</math></li> <li>• Extract <math>\alpha_{CEs}</math> from re-fitting <math>I_{CK}</math> at each <math>T</math> with <math>V_{CEs}(T)</math> as a parameter.</li> <li>• <math>\alpha_{vs}</math> can be taken from literature or adjusted to <math>I_{CK}</math> data</li> </ul>
Extracted (specific) parameters	$\alpha_{vs}, \alpha_{CEs}$
HICUM parameters	$\alpha_{vs}, \alpha_{CEs}$

#### 9.6.19.5 BC breakdown

Accurate determination of the relative temperature coefficients for weak-avalanche current is reflected in modeling characteristics of collector and base currents at different temperatures. Using the simple model equations (8.124) and (8.125), fairly good fit is obtainable.

Data	$I_B(V_{CB} \text{ or } V_{CE})$ data for different temperatures $T$
------	--



Model parameters	none
Required geometry data	none
Procedure:	<ul style="list-style-type: none"> <li>• Determine the avalanche current <math>I_{AVL}</math> from <math>I_B</math> for each <math>T</math></li> <li>• With <math>\alpha_{fav}</math> and <math>\alpha_{qav}</math> as variables, perform a nonlinear optimization on the measured data for <math>I_{AVL}</math> by exercising the compact model with fixed values for <math>f_{AVL}</math> and <math>q_{AVL}</math>, that were determined at the reference temperature <math>T_0</math>.</li> </ul>
Extracted (specific) parameters	$\alpha_{fav}$ , $\alpha_{qav}$
Related HICUM parameters	$\alpha_{fav}$ , $\alpha_{qav}$

The final results of thermal modeling are shown in Fig. 9.19 by comparing the device simulation data with that of HICUM/L2, which visualizes the thermal behavior of the bias dependent collector current and cut-off frequency. A wide temperature range (from 200K to 500K) is considered. A temperature increase shifts the current density towards lower turn-on voltages (cf. left of Fig. 9.19), and the drop of  $f_T$  (cf. right of Fig. 9.19) to lower current densities. The excellent agreement with device simulation data showcases both the thermal modeling capability of HICUM/L2 and the suitability of the extraction techniques presented above. Further verifications for experimental data, including noise and distortion, will be presented in sec. 11.2.

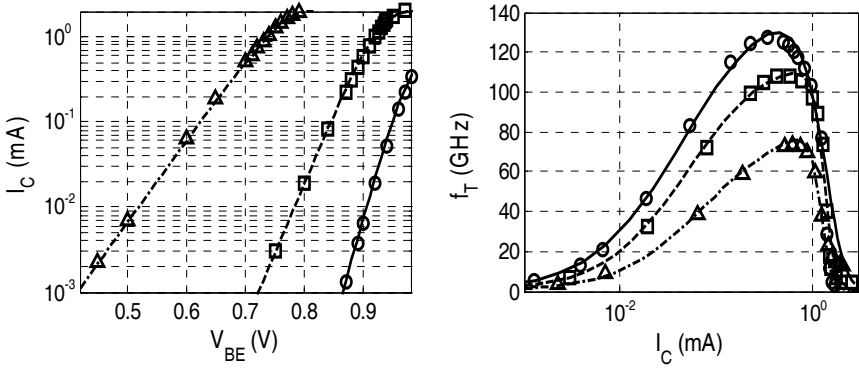


Fig. 9.19: Collector current (left) and cut-off frequency (right) for  $b_{E0}/\mu m = 0.3$  at different temperature  $T/K = 200, 300$ , and  $500$  at  $V_{BC} = 0V$ . Symbols for device simulation and lines for HICUM/L2.

Reliable circuit design requires fabrication tolerances to be taken into account in the model. The PBSA addresses this issue by (i) equations that relate compact model parameters to design rules and (ii) the use of specific electrical parameters and of PCMs [7, 38]. Examples for such PCMs are the internal base sheet resistance and the area-specific internal BC depletion capacitance. A full set of compact model parameters can then be generated for any combination of changes in the PCMs, representing a wafer with different process conditions and, therefore, facilitating parametric yield and also device mismatch simulation. These capabilities depend on the physical limitations of the model. The accuracy is also limited by the simplifications made in deriving such analytical relations. The respective HICUM/L2 relations are quite sophisticated and accurate due to the physical nature of the model. A detailed description of the equations and the methodology is beyond the scope of this book, but sec. 11.2.4 provides an overview on statistical modeling methodologies.

## 9.7 Summary and discussions

The desired deployment of the presented approach is to provide only the specific electrical parameters and design rules rather than many libraries. This situation is similar to that in MOS circuit design, where, however, the

geometry equations are usually included in the simulator's model code. In contrast, the approach presented above uses a separate program that can be adapted quickly to other technologies and models.

It is felt that in order to apply the presented methodology successfully, the employed compact model has to be at least reasonably physical. The basic principle is independent of process scalability. Considering carefully the overall picture, including circuit design and parameter extraction requirements, the methodology readily provides following advantages:

1. Once the parameter extraction is completed, the flexibility of the method facilitates a quick generation of a large number of transistor configurations. As soon as more than about five different transistor configurations are required, which is usually the case, the new method becomes significantly more efficient. In particular, parameters for devices not existing on the (test) chip can be generated at any time prior to and during process development, supporting concurrent design. As a result, design cycle time and time to market can be significantly reduced.
2. A unified set of test structures has been employed that can be used throughout the industry (and has already been available at various companies for many years). A compilation of this experience is contained in [16]. As a result, process selections from an application point of view can be simplified, and nonstandard models with improved capabilities can be used.
3. Parameter determination actually becomes faster and more accurate, since the methodology minimizes non-physical parameter values, which are often obtained from the conventional method of fitting just a single transistor.
4. Applying specific electrical parameters of the different transistor regions (e.g., bottom or perimeter of a junction) allows the construction of more accurate transistor models, e.g., for distributed effects (cf. HICUM/L4). In the conventional case, the parameters of different regions are difficult to extract and require simplified models, in which the elements that represent the different regions have to be lumped together, leading to a loss of scalability.

5. Since the approach is physics-based and linked to PCMs, it facilitates a prediction of the influence of process changes or tolerances on device and circuit behavior.

It is recommended that circuit designers should include separately at least one typical transistor in high-frequency probe pads and, at minimum, the corresponding de-embedding (open) structure in their layout. Together with existing PCMs, this provides for HF circuits that do not function as simulated a way to track down those problems that related to process variations and/or modeling inaccuracies. This is an important contribution to closing the loop between circuit simulation and measurement. The PBSA also works for III/V HBTs as demonstrated in [39].

Although the PBSA is generally preferable, there are occasions when only a single transistor is available for which parameters need to be determined. A corresponding method was given in [26] and the parameters should be extracted in an automated way as in [40]. However, for sophisticated models such as HICUM/L2 it is difficult to find a unique set of parameters due to the fact that the separation of physical effects described by the equivalent circuit is impossible with a single structure. In this case, it is recommended to use a simplified model such as HICUM/L0 which enables a straightforward single-transistor extraction [41].

## 9.8 References

- [1] H.-M. Rein and M. Moeller, "Design considerations for very-high-speed Si-bipolar IC's operating up to 50 Gb/s," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1076–1090, 1996.
- [2] C. D. Hull, private communication. see also:  
C. Hull, S. Kishore, "BiCMOS device modeling for RF IC design", *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 17–22, 1999.
- [3] J. Long, "SiGe radio frequency ICs for low-power portable communication", *Proc. IEEE*, Vol. 93, pp. 1598–1612, 2005.
- [4] T. Massier, H. Graeb, U. Schlichtmann, "The sizing rules method for CMOS and Bipolar analog integrated circuit synthesis", *IEEE Trans. Comp. Aided Des. Integr. Circ. Syst.*, vol. 27, pp. 2209–2222, 2008.

- [5] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener, and A. Koldehoff, "Physics- and Process-Based Bipolar Transistor Modeling for Integrated Circuit Design", IEEE J. Solid-State Circuits, vol. 34, pp. 1136-1149, 1999. see also: Manual Version 5.2, 2006.
- [6] Y. Zimmermann, K. Moebus, H. Wittkopf, M. Schröter, "TRADICA - An integrated modeling tool linking process and circuit design", IEEE SCD, 4 pages, 2009.
- [7] M. Schröter, "Bipolar transistor modeling for high-speed circuit design and process development," in Short Course, IEEE Bipolar Circuits and Technology Meeting, Minneapolis, MN, 1996.
- [8] M. Schröter and H.-M. Rein, "Investigation of very fast and high-current transients in digital bipolar IC's using both a new compact model and a device simulator," IEEE J. Solid-State Circuits, vol. 30, pp. 551-562, 1996.
- [9] B. Ardouin, T. Zimmer, D. Berger, D. Celi, H. Mnif, T. Burdeau, "Transit time parameter extraction for the HICUM bipolar compact model", Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting, pp. 106-109, 2001.
- [10] M. Malorny, M. Schröter, D. Celi, D. Berger, "An improved method for determining the transit time of Si/SiGe bipolar transistors", Proc. Bipolar/BiCMOS Circuits and Technology Meeting, pp. 229-232, 2003.
- [11] HICUM Workshop presentations from 2001 to 2009 at [http://www.iee.et.tu-dresden.de/iee/eb/hic\\_new/hic\\_events.html](http://www.iee.et.tu-dresden.de/iee/eb/hic_new/hic_events.html)
- [12] <http://www.xmodtech.com/Index.html>
- [13] D. Berger, N. Gambette, D. Celi, C. Dufaza, "Extraction of the base-collector capacitance splitting along the base resistance using HF measurements", Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting, pp. 180-183, 2002.
- [14] M. Schröter, "Transient and small-signal high-frequency simulation of numerical device models embedded in an external circuit", COMPEL, Vol. 10, No. 4, pp. 377-378, 1991.
- [15] H.-M. Rein and M. Schröter, "Experimental determination of the internal base sheet resistance of bipolar transistors under forward-bias conditions," Solid-State Electron., vol. 34, pp. 301-308, 1991.

- [16] M. Schröter, S. Lehmann, "The rectangular bipolar transistor tet-rode structure and its application", Proc. ICMTS, Tokyo, Japan, pp. 206-209, 2007.
- [17] M. Schröter, J. Krause, K. Aufinger, B. Ardouin, "Test structures for SiGe HBT compact modeling, parameter extraction and circuit design", Report D4.3.1, DOTFIVE, 2008.
- [18] IC-CAP Modeling Reference: Measurement, Modeling and Simulation of Electronic Components and Circuits, Agilent Eesof EDA, Agilent Technologies, 2000.
- [19] H.K. Gummel, "On the definition of the cutoff frequency  $f_T$ ", Proc. IEEE, Vol. 57, pp. 2159, 1969.
- [20] Mac E. Van Valkenburg, "Network Analysis", Prentice Hall of India Pvt. Ltd., 3rd Edition, 2007.
- [21] H. J. Eul and B. Schiek, "Thru-Match-Reflect: One Result of a Rigorous Theory for De-Embedding and Network Analyzer Calibration," Proc. EuMC, pp. 909-914, 1988.
- [22] E. P. Vandamme, D. Schreurs, G. van Dinther, "Improved three-step de-embedding method to accurately account for the influence of pad parasitics in silicon on-wafer RF test structures," IEEE Trans. Electron Devices, vol. 48, pp. 737-742, 2001.
- [23] H. Cho and D. Burk, "A three step method for the de-embedding of high frequency S parameter measurements," IEEE Trans. Electron Devices, vol. 38, pp. 1371-1375, 1991.
- [24] T. E. Kolding, "Shield-based microwave on-wafer device measurements", IEEE Trans. Microw. Theory Tech., Vol. 49, pp 1039-1044, 2001.
- [25] B. C. Bouma and A. C. Roelofs, "An experimental determination of the forward-biased emitter-base capacitance," Solid-State Electron., vol. 21, pp. 833-836, 1978.
- [26] D. Berger, D. Celi, M. Schröter, M. malorny, T. Zimmer, B. Ardouin, "HICUM parameter extraction methodology for a single transistor geometry", Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting, pp. 116-119, 2002.
- [27] R. Gabl, M. Reisch, M. Pohl, "Improved extraction method for the emitter resistance of bipolar transistors", Proc. Bipolar/BiCMOS Circuits and Technology Meeting, pp. 211-214, 1998.

- [28] H.-M. Rein, "A simple method for separation of the internal and external (peripheral) currents of bipolar transistors," *Solid-State Electron.*, vol. 27, pp. 625–632, 1984.
- [29] H. Q. Tran, M. Schröter, D. J. Walkey, D. Marchesan, T. J. Smy, "Simultaneous extraction of thermal and emitter series resistances in bipolar transistors," *Proc. IEEE Bipolar Circuits and Technology Meeting*, Minneapolis, MN, 1997, pp. 170–173.
- [30] D.J. Walkey, T.J. Smy, D. Marchesan, H. Tran, C. Reimer, T.C. Kleckner, M.K. Jackson, M. Schröter and J.R. Long, "Extraction and modeling of thermal behavior in trench isolated bipolar structures", *IEEE Bipolar and BiCMOS Circuits and Technology Meeting*, Minneapolis, pp. 97-100, 1999.
- [31] T. Vanhoucke, H. M. J. Boots, W. D. van Noort, "Revised Method for Extraction of the Thermal Resistance Applied to Bulk and SOI SiGe HBTs", *IEEE Electron Dev., Lett.*, Vol. 25, pp. 150- 152, 2004.
- [32] H.-M. Rein, "Proper choice of the measuring frequency for determining  $f_T$  of bipolar transistors," *Solid-State Electronics*, vol. 26, pp. 75–83 and p. 929, 1983.
- [33] M. Schröter, Z. Yan, T.-Y Lee, W. Shi, "A compact tunneling current and collector breakdown model," *Proc. IEEE Bipolar Circuits and Technology Meeting*, Minneapolis, 1998, pp. 203–206.
- [34] M. Pfost, H.-M. Rein, and T. Holzwarth, "Modeling substrate effects in the design of high-speed Si-bipolar IC's," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1493–1501, 1996.
- [35] M. Pfost and H.-M. Rein, "Modeling and measurement of substrate coupling in Si-bipolar IC's up to 40 GHz," *IEEE J. Solid-State Circuits*, vol. 34, pp. 582–591, 1998.
- [36] S. Harker, R. Havens, J. Paaschens, D. Szmyd, L. Tiemeijer, "An S-parameter technique for substrate resistance characterization of bipolar transistors", *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, Minneapolis, pp. 176-179, 2000.
- [37] U. Basaran, M. Berroth, "An accurate method to determine the substrate network elements and base resistance", *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, Minneapolis, pp. 93-96, 2003.

- [38] M Schröter, H. Wittkopf, W. Kraus, “Statistical modeling of bipolar transistors”, Proc. Bipolar Circuits and Technology Meeting (BCTM), pp. 54-61, 2005.
- [39] Y. Zimmermann, M. Schröter, P. Zampardi, A. Klimashov, G. Tkachenko, “III-V HBT Modeling, scaling and parameter extraction using TRADICA and HICUM”, IEEE Workshop on Power Amplifiers, San Diego, CA, 2003.
- [40] E. Castro, S. Coco, A. Laudani, L. Lo Nigro, and G. Pollicinodes, “A New Tool for Bipolar Transistor Characterization Based on HICUM”, SIMAI Congress, 7 pages, 2007.
- [41] S. Fregonese, S. Lehmann, T. Zimmer, M. Schröter, D. Celi, B. Ardouin, H. Beckrich, P. Brenner, W. Kraus, “A Computationally Efficient Physics-Based Compact Bipolar Transistor Model for Circuit Design—Part II: Parameter Extraction and Experimental Results”, IEEE Trans. Electron Dev., Vol. 53, pp. 287- 295, 2006.



# **Chapter 10**

## **Model Hierarchy**

## 10.1 Introduction

The complexity of compact models can vary widely depending on the application focus, but generally has increased significantly over time as a result of an ever increasing complexity of device design and associated number of physical effects. For supporting efficient circuit design a model hierarchy with three significantly different levels of model complexity has been developed. Existing advanced models for bipolar transistors such as HICUM/L2 [1] contain a representation of all known *relevant* physical effects. Although the latter are described already in simplified form, the resulting equivalent circuit and equations appear to be fairly complicated to model users. However, at least during the initial circuit design phase, it is usually advantageous to start with a simple transistor model that can be quickly reduced to just the essential transistor features and is easily understandable for designers. This facilitates a fast evaluation of the basic circuit functionality, before spending significant time on longer optimization cycles, using more accurate models. Also, being able to start with a simple model and then gradually move to more sophisticated models provides a better feeling for the impact of physical effects on circuit characteristics. While integrated circuit design and optimization generally requires geometry scalable models, sometimes discrete devices have to be dealt with. Parameter determination from such a “single geometry transistor” naturally is quite difficult for sophisticated models, unless they are simplified with respect to, e.g., geometry effects. Another example is a parasitic transistor, e.g. in CMOS processes, that requires just a reasonably accurate description of the fundamental bipolar transistor action.

As a consequence, the simplified model HICUM/L0 has been formulated more recently. The goal of this development is to provide a model with similar simplicity as the legacy SGPM but to eliminate as many as possible issues of the SGPM for advanced processes *without increasing the computational and parameter extraction effort*. The simplicity of HICUM/L0 along with the existing HICUM/L2 infrastructure for simulator implementation and parameter extraction have led to quite a quick adoption by industry. At the time this book was written, the version 1.2 of HICUM/L0 became available in commercial simulators and first foundry PDKs.

The final verification of circuit behavior eventually requires HICUM/L2 or an even more sophisticated model, at least for the most critical transistors and signal paths. For instance, transistors in driver circuits and power amplifiers (PAs) tend to become quite large and often are realized as multi-finger structures or are arranged as arrays of fixed-finger transistor basic cells. The size of the overall device may lead to distributed electrical and thermal effects. The latter result from self-heating and the mutual coupling between fingers and cells [2]. Also, operation near collector breakdown can cause a three-dimensional pinch-in effect [3]. Obviously, lumped models cannot capture such distributed effects. Therefore, HICUM/L4 has been developed to allow a suitable partitioning of a transistor structure into a distributed model. The goal is to provide an easy path from a conventional process-based scalable lumped model, such as HICUM/L2, to a model that covers distributed effects and helps circuit designers to detect pitfalls in the verification or critical phase of their design.

Figure 10.1 shows the resulting model hierarchy and its relation to the SPICE Gummel Poon model (SGPM). Here, SGPM/L0 corresponds to the original and most widely implemented version, while SGPM/L2 corresponds to an extended EC built around the SGPM/L0 as core in order to better capture the characteristics of advanced technologies.

The hierarchy is centered around HICUM/L2, which contains the most advanced formulations of these models. Therefore, its physical basis and scalability allows users to *generate without any additional parameter extraction effort* the other models (and their parameters). Since the relevant process-specific parameters and layout dimensions are already available in TRADICA (cf. [4] and ch. 11), the latter allows users to very efficiently realize a self-consistent model hierarchy that requires very low maintenance for a foundry. Integration of TRADICA in a design system enables users to quickly adapt the employed model(s) to the given design task. Obviously, the need to flexibly change the EC topology makes such a hierarchy generation difficult to realize with just simulator preprocessor scripts within a model. The remainder of this chapter provides a detailed derivation and description of HICUM/L0 and a rough overview on the concept of HICUM/L4.

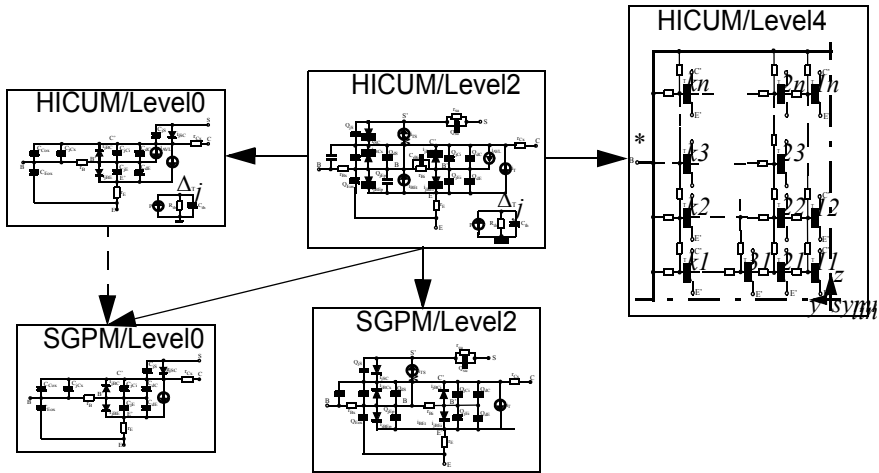


Fig. 10.1: Model hierarchy centered around HICUM/L2. Parameters of the other models can be generated from HICUM/L2 parameters and characteristics.

## 10.2 HICUM/L0: a simplified model

HICUM/Level0 [5]-[10] has been developed fairly recently and has the same number of *electrical* nodes as the standard SGPM (Level0), but significantly improved model equations. The latter have been mostly borrowed from HICUM/L2 and have then partially been reformulated in a simplified form. They include first-order effects such as minority charge storage at medium and high injection, BC avalanche breakdown, self-heating, a consistent bias dependent description of the internal base resistance, and collector punch-through in the BC depletion charge. As a result, of the improved and more physical model equations compared to the SGPM, also parameter extraction difficulties are reduced.

Besides enabling designers to obtain a feeling for circuit behavior by performing simple calculations, the main purpose of a simplified model is to speed up simulation in general (especially for large circuits) and also to meet today's requirements for advanced integrated circuit design. This includes facilitating an easy migration from a conventional, single-transistor-based, to a process-based scalable parameter extraction and model usage. The major disadvantage of a model simplification is a reduction of

the validity range, caused by either inaccurate equations or completely missing physical effects. Nevertheless, the obtained results can always be verified against the more sophisticated levels (2 and 4) of HICUM. Despite the simplifications, HICUM/L0 still fulfils the following requirements:

- Its parameters can be generated from HICUM/L2 avoiding users to add any extra burden in terms of parameter extraction. For instance, a L0 library can be generated either from TRADICA [11], partially using directly process specific data, or within a simulator from a L2 parameter set and an extraction preprocessor.
- HICUM/L0 keeps a similarity to the standard SGPM, which most circuit designers are familiar with.
- A clear definition of the model parameters and their physical background is given, providing modeling and design engineers with the often missing, but required, information on fundamental model limitations. This allows users, at least to a certain extent, to set up predictive and statistical design for larger circuits.

HICUM/L0 is made available through Verilog-A code [6], which can be exercised with and compiled into commercial circuit simulators (e.g. ([13, 14]).

The subsequent derivations and discussions give a rough idea of the amount of effort that needs to be spent to create a *working* simplified model, which has often been requested by users (and not only for bipolar transistors). In contrary to the widespread perception, such a model is *not* obtained by just dropping terms and elements of a fully physical model version. As will be seen, there are several important cases in which a new expression needs to be derived and extensively tested.

### 10.2.1 Large-signal model formulation

In this section the HICUM/L0 equivalent circuit (EC) and model equations are derived from HICUM/L2. This provides a clearly defined relation between L0 and L2 model parameters and also an overview on the expected limitations of L0. Thus, it is assumed that the reader is familiar with the formulations of HICUM/L2.

### 10.2.1.1 Equivalent circuit

The large-signal EC is shown in Fig. 10.2. It results from the HICUM/L2 EC by merging elements and neglecting certain physical effects, resulting in basically the same EC structure as the standard SGPM, except for the self-heating network. In contrast to HICUM/L2, the various internal, peripheral and external transistor regions cannot be clearly distinguished any more in the EC. The loss of direct physical correspondence is a result of the simplification. Compared with the HICUM/L2 EC, the BE tunneling current, the lateral NQS effect, the barrier related base current recombination, and the substrate coupling network (which can always be added per sub-circuit), have been removed from the HICUM/L0 EC. All junctions are represented by a diode since this often aids in the DC convergence behavior in a circuit simulator. In addition to the simplification of the EC mentioned above, vertical NQS effects and collector current spreading have also been neglected within the model equations, and some other effects are only included in simplified form.

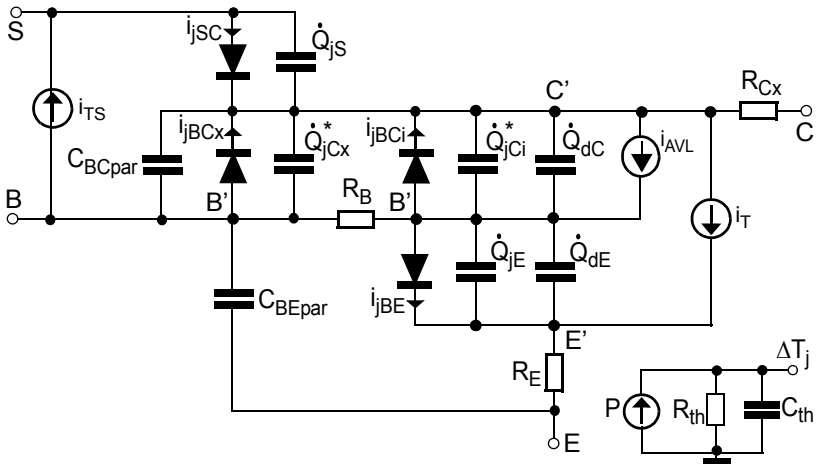


Fig. 10.2: Large-signal equivalent circuit of HICUM/L0 for an npn transistor. It is recommended to make the thermal node  $\Delta T_j$  and the internal base node B' accessible to the outside in a simulator implementation in order to allow modeling of distributed thermal and electrical effects.

### 10.2.1.2 Depletion charges and capacitances

Modeling of the bias dependence of the depletion charges ( $Q_j$ ) and capacitances ( $C_j$ ) is based on the well-proven equations of HICUM/L2 described in section 8.1.1. The mapping equations for the merged elements and differences to HICUM/L2 are discussed below.

#### A Base-emitter junction

Compared to HICUM/L2, the BE depletion capacitance components of the *internal and peripheral* transistor are merged into a single element,

$$C_{jE} = C_{jEi} + C_{jEp}, \quad (10.1)$$

using merged values for the zero-bias capacitance

$$C_{jE0} = C_{jEi0} + C_{jEp0}, \quad (10.2)$$

which is directly related to the respective HICUM/L2 values for the internal and peripheral BE junction. Like for HICUM/L2, the model avoids the numerical overflow at the built-in voltage by smoothly limiting the capacitance to its (measurable) maximum value  $a_{jE}C_{jE0}$ . Hence, the total capacitance consists of a classical portion and a component for high forward bias as in (8.2),

$$C_{jE} = \frac{C_{jE0}}{(1 - v_j/V_{DE})^{z_E}} \cdot \frac{dv_j}{dv_{B'E'}} + a_{jE}C_{jE0} \left(1 - \frac{dv_j}{dv_{B'E'}}\right). \quad (10.3)$$

The smoothing function  $v_j$  for the auxiliary voltage, its derivative, and the voltage  $V_f$  at which the forward bias capacitance of the classical expression intercepts the maximum constant value are given by (8.3) to (8.5). The value of  $a_{jE}$  is given by  $C_{jE,max} = C_{jEi,max} + C_{jEp,max}$ , and can be related to HICUM/L2 parameters:

$$a_{jE} = \frac{a_{jEi}C_{jEi0} + a_{jEp}C_{jEp0}}{C_{jE0}}. \quad (10.4)$$

The corresponding charge equation follows (8.1) and reads as

$$Q_{jE} = \frac{C_{jE0}V_{DE}}{1 - z_E} \left[ 1 - \left(1 - \frac{v_j}{V_{DE}}\right)^{(1 - z_E)} \right] + a_{jE}C_{jE0}(v_{B'E'} - v_j). \quad (10.5)$$

For each transistor configuration, the values of  $V_{DE}$  and  $z_E$  can be extracted from measurements or by exercising HICUM/L2 equations at the appropriate forward bias operating points.

## B Base-collector junction

Due to the simplified EC, the BC depletion capacitance must now be partitioned across the *total* base resistance. The resulting “internal” and “external” charge portions have been labelled in Fig. 10.2 by a “\*” in order to distinguish them from the physical components (and values) used in HICUM/L2. For transistors with selectively implanted collectors, which is the standard option in most processes, the (original) physical components  $C_{jCi}$  and  $C_{jCx}$ , as they are used in HICUM/L2, differ in their bias dependence, particularly in terms of punch-through behavior. Therefore and in order to provide sufficient flexibility for modeling the high-frequency characteristics, a separate set of model parameters has been maintained for the elements  $Q_{jCi}^*$  and  $Q_{jCx}^*$ . In addition, the following criteria should be met by the simplified model:

- It should be possible to extract the model parameters from single transistors, i.e. without the knowledge of geometry specific data and, thus, without the option of distinguishing between  $C_{jCi}$  and  $C_{jCx}$ .
- The partitioning of the capacitances should be as simple as possible and requiring only a minimum number of model parameters.
- The formulation should also allow the users to calculate HICUM/L0 parameters directly from a geometry scalable methodology used for generating HICUM/L2 parameters.

The goal therefore is to formulate and implement a flexible scheme that allows users to match all of the above requirements. This has been done by implementing the BC depletion capacitance as shown in Fig. 10.3a. The “internal” depletion portion,  $C_{jCi}$ , maintains its separate parameter set. The external depletion capacitance  $C_{jCx} = C_{jCx1} + C_{jCx2}$  (and corresponding charge  $Q_{jCx}$ ) consists of two elements that are split across  $R_B$  according to a partitioning factor  $f_{BC}$  defined by the user. Both  $C_{jCx1}$  and  $C_{jCx2}$  use the same voltage dependence but a different zero-bias value. The corresponding code implementation is shown in Fig. 10.3b.



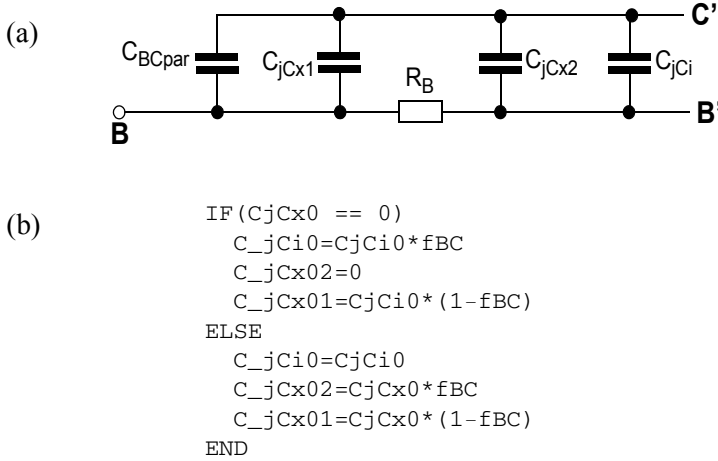


Fig. 10.3: Implementation of a flexible partitioning scheme for the various BC capacitance components: (a) equivalent circuit; (b) code example.

This implementation not only permits modeling of the internal and external depletion capacitance with *separate* parameters (if available) but also provides a simple scheme for partitioning the total capacitance arbitrarily across  $R_B$  via adjusting  $f_{BC}$ . Note, that the total capacitance will only be correct *exactly* for zero-bias due to the different values for  $V_{BC'}$  and  $V_{B'C'}$ . The use of this partitioning scheme is illustrated briefly below for the two possible cases of parameter extraction.

- If a geometry scalable extraction has been performed, the parameters  $C_{jCi0}$  and  $C_{jCx0}$  are known and specified by the user as non-zero values. Also, the remaining parameters for each capacitance are usually different. In this case, only the external capacitance is partitioned into  $C_{jCx1}$  and  $C_{jCx2}$  according to  $f_{BC}$ , which is equivalent to the parameter  $X_{JC}$  in the SGPM; e.g.,  $f_{BC} = 0$  corresponds to the case in which the entire external BC depletion capacitance is located between the nodes B and C'.
- If a single device is considered, only the parameters for the *total* depletion capacitance can be determined. Therefore, the parameter list contains  $C_{jCi0} > 0$  and  $C_{jCx0} = 0$ , while the remaining parameters of the external portion are set equal to the ones specified for  $C_{jCi}$ . In this case, the total capacitance is partitioned according to  $f_{BC}$  and the element  $C_{jCx2}$  is omitted.

The bias dependent BC depletion capacitance formulation for HICUM/L0 follows the HICUM/L2 formulation described in ch 8.1.1. The equations include collector punch-through and also avoid the numerical overflow at the built-in voltage. To take into account punch-through for the internal and external junction, the model parameters  $V_{PTCi}$  (cf. (8.10)) and  $V_{PTCx}$ , respectively, need to be specified.

### C Collector-substrate junction

The element  $Q_{jS}$  in the EC of Fig. 10.2 represents the *total* collector-substrate depletion charge. Its voltage dependence and that of the associated capacitance  $C_{jS}$  are modelled the same way as  $Q_{jCx}$  and  $C_{jCx}$ . In most processes, punch-through is not relevant and the respective parameter can be omitted. The parameters can be either determined from measured capacitance-voltage data or taken from HICUM/L2.

#### 10.2.1.3 Minority charge

The transit time and minority charge expressions results from a simplification of the HICUM/L2 equations discussed in sec. 8.1.2. The formulation of the forward minority charge  $Q_f$  is strongly based on an accurate description of the transit time  $\tau_f$ . As in HICUM/L2 the bias dependence of  $\tau_f$  is described by

$$\tau_f = \tau_{f0}(v_{B'C'}) + \Delta\tau_f(i_{TF} v_{C'E'}), \quad (10.6)$$

with the low current component  $\tau_{f0}$ , the high-current component  $\Delta\tau_f$  and  $i_{TF}$  as the forward transfer current.

The low-current transit time is modeled the same way as in HICUM/L2 and reads (cf. (8.25)):

$$\tau_{f0}(v_{B'C'}) = \tau_0 + \Delta\tau_{0h}(c - 1) + \tau_{Bfvl}\left(\frac{1}{c} - 1\right). \quad (10.7)$$

The resulting voltage function  $c = C_{jCi0}^*/C_{jCi}^* \approx C_{jCi0}/C_{jCi}$  does not depend on the zero-bias capacitance. The respective forward minority charge is given by  $Q_{f0} = \tau_{f0}i_{TF}$ .

Compared to HICUM/L2 the current dependence of  $\Delta\tau_f$  is simplified to the formulation (8.48), including base, collector and emitter component:

$$\Delta\tau_f = \tau_{hes} w^2 \left[ 1 + \frac{2I_{CK}}{i_{Tf} \sqrt{i^2 + a_{hc}}} \right] + \tau_{E\theta} \left( \frac{i_{Tf}}{I_{CK}} \right)^{g_{\tau E}} \quad (10.8)$$

with

$$i = 1 - \frac{I_{CK}}{i_{Tf}} \quad (10.9)$$

and the normalized collector injection width

$$w = \frac{i + \sqrt{i^2 + a_{hc}}}{1 + \sqrt{1 + a_{hc}}} \quad (10.10)$$

This formulation still hinges on the physics-based critical current  $I_{CK}$  as key variable, which determines the voltage dependent shift and is described by (8.27) and (8.28). The associated model parameters  $r_{Ci0}$ ,  $V_{lim}$ ,  $V_{PT}$  and  $V_{CEs}$  are also used in HICUM/L0. Above formulation neglects the bias dependent portion of the collector current spreading formulation and the BC barrier effect, thus allowing us to merge the base and collector component of  $\Delta\tau_f$  into a single expression.

The total forward minority charge  $Q_f$  employed for dynamic transistor operation is then obtained analytically by integrating  $\tau_f$  over  $i_{Tf}$

$$Q_f = \tau_{f0} i_{Tf} + \Delta Q_f,$$

with the high-current contribution given by

$$\Delta Q_f = \tau_{hes} i_{Tf} w^2 + \tau_{E\theta} \left( \frac{i_{Tf}}{I_{CK}} \right)^{g_{\tau E}} \frac{i_{Tf}}{1 + g_{\tau E}} \quad (10.11)$$

In order to maintain an accurate physics-based description of the transit time as function of  $v_{CE}$ , temperature, doping and geometry, the bias independent portion of the collector current spreading factor  $f_{cs}$  is included in  $I_{CK}$  according to the discussion in sec. 8.2.1. By adjusting the parameter  $\tau_{hes}$  properly, the impact of possible current spreading on  $\Delta\tau_f$  can still be described reasonably well up to the validity limit of HICUM/L0. The reverse minority charge is simply  $Q_r = \tau_r i_{Tr}$  with the reverse transit time  $\tau_r$  as model parameter. The formulations for the corresponding diffusion charges are similar to (8.23) and (8.51), respectively.

Above model equations facilitate an easy extraction of parameters, particularly from single devices at the expense of limited accuracy at high current densities and geometry scalability. The latter can still be achieved though by extracting the parameters from the full HICUM/L2 formulation.

#### 10.2.1.4 Quasi-static transfer current

In contrast to the implicit formulation of the GICCR, the goal here is to obtain an *explicit* expression for the transfer current that is simple but accurate enough for the practically most widely utilized bias region and allows hand or spreadsheet calculations. The only known true explicit formulation has been the one implemented in the SGPM, but has been proven inadequate for modern technologies (cf. sec. 4.3.1.2). An often made request that goes even further is to separate the DC from the charge description in the (G)ICCR. Both approaches will definitely come at the expense of losing - to a significant extent - the physical basis of the transfer current description. An explicit formulation addressing the first issue mentioned above was derived in [7] and has been demonstrated to yield significantly improved results compared to the SGPM at higher current densities. The industrial deployment of the corresponding HICUM/L0 version 1.12 exposed some weaknesses of this transfer current formulation that have been eliminated more recently and also address the second request mentioned earlier. The resulting simplified transfer current formulation, which completely decouples DC and AC description, is derived below.

According to (4.52) to (4.55), (4.61) and (4.63a, b) the exact solution of the transfer current can be written as

$$i_T = i_{Tf} - i_{Tr} = I_S \frac{\exp\left(\frac{v_{B'E}}{V_T}\right) - \exp\left(\frac{v_{B'C}}{V_T}\right)}{q_{pT}} \quad (10.12)$$

with the saturation current

$$I_S = q A_E c_0 / Q_{p0h,0}, \quad (10.13)$$

$Q_{p0h,0}$  as the bias independent weighted zero-bias hole charge, and the normalized weighted hole charge

$$q_{pT} = 1 + \frac{\Delta Q_{p0h}}{Q_{p0h,0}} + \frac{Q_{jEih}}{Q_{p0h,0}} + \frac{Q_{jCih}}{Q_{p0h,0}} + \frac{\Delta Q_{mh}}{Q_{p0h,0}}, \quad (10.14)$$

which is often referred to as normalized *base* charge. The second term on the r.h.s. represents the bias dependent portion  $\Delta Q_{p0h}$  of the weighted zero-bias hole charge  $Q_{p0h,0}$  (cf. Fig. 4.20), which is close to zero for BJTs. This bias dependence results from weight function changes close to the SCR edges. It allows  $\Delta Q_{p0h} = \Delta Q_{p0h,jE} + \Delta Q_{p0h,jC}$  to be formally distributed to the modified weighted depletion charges

$$Q_{jEid} = Q_{jEih} + \Delta Q_{p0h,jE}, \quad Q_{jCid} = Q_{jCih} + \Delta Q_{p0h,jC}, \quad (10.15)$$

changing their bias dependence. In principle, this could still be described by the product of a weight factor and the actual internal depletion charge. However, the necessary bias dependent description of the weight factor, which is different from HICUM/L2, would be too complicated for a simplified model. Furthermore, due to the simplification of the HICUM/L0 equivalent circuit the internal depletion capacitances and charges are not available. Therefore, the modified weighted depletion charges are approximated directly by the BE and BC depletion charge *formulations*, but with a different set of parameters compared with the actual charges. In other words, the various bias dependences are lumped into a simple depletion charge formulation but with its parameters *determined from the static transfer current characteristic* rather than from small-signal capacitance measurements. This idea was successfully exercised in [15]. In order to avoid having to determine capacitance values, the charge expressions need to be normalized, preferably to a zero-bias capacitance. The resulting consequences are discussed next.

The normalized BE depletion term can be written as

$$\frac{Q_{jEid}}{Q_{p0h,0}} = \frac{Q_{jEid}/C_{jEi0}}{Q_{p0h,0}/C_{jEi0}} = \frac{v_{jEid}}{V_{Er}} \quad (10.16)$$

with  $v_{jEid}$  as charge related voltage function. The model parameter

$$V_{Er} = \frac{Q_{p0h,0}}{C_{jEi0}} \approx \frac{Q_{p0}}{h_{jei} C_{jEi0}}, \quad (10.17)$$

is referred to as a reverse Early voltage in SGPM or VBIC [16] since it determines the inverse output conductance at low injection. Obviously,  $V_{Er}$  is a derived or composite parameter that contains different basic physical variables and effects. For instance it includes the GICCR weight factor. The rightmost expression in (10.17) can be used to estimate  $V_{Er}$  from HICUM/L2 parameters. The default values for the parameters ( $V_{DEid}$ ,  $z_{Eid}$ ,  $a_{jEid}$ ) of the voltage function  $v_{jEid}$  can be set to the values of the actual depletion capacitance. Since generally similar accuracy for  $Q_{jEi}$  can be obtained with different value pairs for ( $V_{DEi}$ ,  $z_{Ei}$ ), it is recommended to keep  $z_{Eid} = z_{Ei}$  and to only use  $V_{DEid}$  and  $a_{jEid}$  as parameter.

In a similar way the normalized BC depletion charge is replaced by the simpler expression

$$\frac{Q_{jCid}}{Q_{p0h,0}} = \frac{Q_{jCi}/C_{jCi0}}{Q_{p0h,0}/C_{jCi0}} = \frac{v_{jCid}}{V_{Ef}} \quad (10.18)$$

with  $v_{jCid}$  as charge related voltage function. The model parameter

$$V_{Ef} = \frac{Q_{p0h,0}}{C_{jCi0}} \approx \frac{Q_{p0}}{h_{jci} C_{jCi0}} \quad (10.19)$$

can also be interpreted as a forward Early-voltage since it determines the output conductance at low injection. Like  $V_{Er}$ ,  $V_{Ef}$  is a derived or composite parameter that contains different basic physical variables.

The approach for simplifying the normalized depletion charges includes the (bias dependent) weight factors. Such a formulation also eliminates the need to know the absolute value of the internal depletion capacitances. However, it is difficult to assign a clear physical meaning to the parameters of the voltage functions  $v_{jEid}$  and  $v_{jCid}$ , and their bias dependence. In order to keep the simplicity of the model and to improve its flexibility for a wide variety of applications the emission factors  $\mathbf{m}_{Cf}$  (and  $\mathbf{m}_{Cr}$ ) are also being added for taking into account, e.g., thermionic emission and tunnelling in III/V HBTs.

Up to this point, the relation for the forward transfer current reads

$$i_{Tf} = \frac{I_S}{1 + \frac{v_{jEid}}{V_{Er}} + \frac{v_{jCid}}{V_{Ef}} + \frac{\Delta Q_{mh}}{Q_{p0h,0}}} \exp\left(\frac{v_{B'E}}{m_{Cf} V_T}\right). \quad (10.20)$$

Note, that the definition of the Early-voltages and the “DC” depletion charge related voltage functions decouple the description of the DC behavior from that of the AC behavior at low injection. This goal now also needs to be achieved at higher injection through proper simplifications of the (normalized) minority charges.

The accurate and physically correct description of  $i_{Tf}$  (and  $i_T$ ) at medium and high injection requires a model for the weighted minority charge  $\Delta Q_{mh}$ , which can be approximated by the sum ( $Q_{f,T} + Q_{r,T}$ ) of a forward and a reverse component that are used in the GICCR of HICUM/L2 (cf. (8.56)) and which reads in normalized form as

$$\frac{Q_{f,T}}{Q_{p0h,0}} = \frac{\tau_{f0}}{Q_{p0h,0}} i_{Tf} + \frac{\Delta Q_{f,T}}{Q_{p0h,0}} = \frac{i_{Tf}}{I_{Qf}(v_{B'C})} + \Delta q_{fT} \quad (10.21)$$

where the knee current of the DC forward transfer curve is defined as

$$I_{Qf}(v_{B'C}) = Q_{p0h,0} / \tau_{f0}(v_{B'C}). \quad (10.22)$$

The voltage dependence is based on that of the low-current transit time and reads

$$I_{Qf}(v_{B'C}) = \frac{I_{Qf}}{1 + f_{iqf}[\tau_{f0}(v_{B'C})/\tau_0 - 1]} \quad (10.23)$$

with the model parameters  $I_{Qf}$  and  $f_{iqf}$ . The latter acts as a flag that allows us to entirely deactivate the voltage dependence. Similarly, the normalized inverse operation related minority charge can be written as

$$\frac{Q_{r,T}}{Q_{p0h,0}} \approx \frac{i_{Tr}}{I_{Qr}} \quad (10.24)$$

with  $I_{Qr}$  as model parameter. Physically, the two currents  $I_{Qf}$  and  $I_{Qr}$  represent base conductivity modulation, if for  $\tau_{f0}$  just the *base* transit time would be inserted.

With the above approximations the normalized total hole charge at *low* current densities can be written as ( $\Delta q_{fT} = 0$ ),

$$q_{pT,l} = 1 + \frac{v_{jEid}}{V_{Er}} + \frac{v_{jCid}}{V_{Ef}} + \frac{i_{Tf}}{I_{Qf}} + \frac{i_{Tr}}{I_{Qr}}. \quad (10.25)$$

Defining ideal current components,

$$i_{Tfi} = I_S \exp\left(\frac{v_{BE}}{m_{Cf} V_T}\right), \quad i_{Tri} = I_S \exp\left(\frac{v_{BC}}{m_{Cr} V_T}\right), \quad (10.26)$$

as well as the current independent normalized depletion related charge

$$q_j = 1 + \frac{v_{jEid}}{V_{Er}} + \frac{v_{jCid}}{V_{Ef}}, \quad (10.27)$$

and the current dependent normalized low-injection minority charge

$$q_{fl} = \frac{i_{Tfi}}{I_{Qf}} + \frac{i_{Tri}}{I_{Qr}} \quad (10.28)$$

leads to the quadratic equation

$$q_{pT,l} = q_j + \frac{q_{fl}}{q_{pT,l}}. \quad (10.29)$$

Its solution gives the normalized *low-injection* hole charge

$$q_{pT,l} = \frac{q_j}{2} + \sqrt{\left(\frac{q_j}{2}\right)^2 + q_{fl}} \quad (10.30)$$

This results in the well-known low-injection approximation of the forward and reverse transfer current components:

$$i_{Tfl} = \frac{i_{Tfi}}{q_{pT,l}} \quad \text{and} \quad i_{Trl} = \frac{i_{Tri}}{q_{pT,l}}. \quad (10.31)$$

At higher current densities the stronger than linear increase of the minority charge needs to be taken into account for accurately modeling the transfer current and transconductance. With the simplifications so far the total normalized hole charge reads with (10.21), (10.24), and (10.27)



$$q_{pT} = q_j + \frac{i_{Tf}}{I_{Qf}(v_{B'C})} + \frac{i_{Tr}}{I_{Qr}} + \Delta q_{fT}. \quad (10.32)$$

The additional weighted forward minority charge contribution is given by  $Q_{m,T} - Q_{f0} - Q_r$  from (8.56). The various components are described by (10.11) after splitting the base and collector distribution according to (8.39) and multiplying with the associated weight factors, leading to

$$\Delta q_{fT} = \frac{[(1 - f_{thc}) + h_{fC}f_{thc}]\tau_{hcs}i_{Tf}w^2 + h_{fE}\tau_{Ef0}\left(\frac{i_{Tf}}{I_{CK}}\right)^{g_{\tau E}} \frac{i_{Tf}}{1 + g_{\tau E}}}{Q_{p0h,0}}.$$

This expression does not permit an explicit solution for  $q_{pT}$ . As a first step towards further simplification,  $g_{\tau E}$  is set to 1. Also, the parameters

$$I_{Qfh} = \frac{Q_{p0h,0}}{[(1 - f_{thc}) + h_{fC}f_{thc}]\tau_{hcs}}, \quad (10.33)$$

which is usually much smaller than  $I_{Qf}$  and

$$t_{fh} = \frac{h_{fE}}{[(1 - f_{thc}) + h_{fC}f_{thc}]2\tau_{hcs}} \frac{\tau_{Ef0}}{I_{CK}} \quad (10.34)$$

are introduced. This reduces  $\Delta q_{fT}$  to

$$\Delta q_{fT} = \left[ w^2(i_{Tf}) + t_{fh} \frac{i_{Tf}}{I_{CK}} \right] \frac{i_{Tf}}{I_{Qfh}}. \quad (10.35)$$

Inserting above expression into (10.32) results in

$$q_{pT} = q_j + \frac{i_{Tf}}{I_{Qf}(v_{B'C})} + \frac{i_{Tr}}{I_{Qr}} + \frac{i_{Tf}w^2(i_{Tf}) + \frac{t_{fh}}{I_{CK}}i_{Tf}^2}{I_{Qfh}}. \quad (10.36)$$

According to (10.12) and (10.26), the transfer current components read

$$i_{Tf} = \frac{i_{Tfi}}{q_{pT}} \quad \text{and} \quad i_{Tr} = \frac{i_{Tri}}{q_{pT}}. \quad (10.37)$$

Inserting these into (10.36) and defining the charge variables

$$\Delta q_{BCfi} = \frac{w^2(q_{pT})}{I_{Qfh}} i_{Tfi}, \quad \Delta q_{Efi} = t_{fh} \frac{i_{Tfi}^2}{I_{CK} I_{Qfh}}, \quad (10.38)$$

gives with (10.28)

$$q_{pT} = q_j + \frac{q_{fl}}{q_{pT}} + \frac{\Delta q_{BCfi}}{q_{pT}} + \frac{\Delta q_{Efi}}{q_{pT}^2}. \quad (10.39)$$

This still leads to the third-order equation

$$q_{pT}^3 - q_j q_{pT}^2 - q_{fl} q_{pT} - \Delta q_{BCfi}(q_{pT}) q_{pT} - \Delta q_{Efi} = 0, \quad (10.40)$$

in which  $\Delta q_{BCfi}$  indirectly depends on  $q_{pT}$ . Therefore, it is desirable to simplify (10.40) further so that it can be solved directly. A first attempt was to evaluate  $w$  in  $\Delta q_{BCfi}$  as function of the already known current  $i_{Tfl}$  and to obtain an explicit expression from the Cardani solution [17] of (10.40). Unfortunately, the corresponding investigation showed that this approach does not *always* yield a *real valued* solution. As the next attempt,  $i_{Tf}$  was replaced by  $i_{Tfl}$  already in (10.35) resulting in

$$q_{pT} = q_{pT,l} + \Delta q_{fT}(i_{Tfl}) \quad (10.41)$$

and the original transfer current formulation [7]

$$i_{Tf} = \frac{i_{Tfl}}{1 + \frac{\Delta q_{fT}(i_{Tfl})}{q_{pT,l}}} \quad \text{and} \quad i_{Tr} = \frac{i_{Trl}}{1 + \frac{\Delta q_{fT}(i_{Tfl})}{q_{pT,l}}}. \quad (10.42)$$

Here,  $\Delta q_{fT}(i_{Tfl})/q_{pT,l}$  is considered as a *high-current correction* of the low-current solution. Within its validity limits excellent results were in fact obtained in [7] with above equation for a variety of different process technologies.

It was reported later on in [9] that, for certain model parameter choices, with (10.42) a negative transconductance  $g_m$  can occur over a certain bias range at high current densities. This can be traced back to the fact that  $i_{Tfl}$  is increasing much faster with  $v_{B'E'}$  than the actual current  $i_{Tf}$ , and so does  $\Delta q_{fT}(i_{Tfl})$ . This in turn leads to a faster increase of the denominator in (10.42) than the numerator for the bias range within which  $w(i_{Tf})$  transitions from its low-injection to its high-injection value. Another issue was

that at *very* high injection the emitter related term causes the transfer current to flatten out and, thus,  $g_m = 0$ . Although it is not clear whether in a SiGe HBT with a strong BC barrier effect  $g_m$  can actually become temporarily zero or negative, and in the original derivation of the simplified model the corresponding (very) high injection region was considered to be outside of the intended validity limits, a possible negative or zero transconductance is not desirable for convergence reasons.

A more reliable approach for solving (10.40) has to be centered around (i) finding a suitable approximation of  $w$  as function of bias in  $\Delta q_{BCfi}$  of (10.38) and (ii) the emitter related contribution  $\Delta q_{Efi}$ , which causes (10.40) to be a third-order equation in the first place. The impact of  $\Delta q_{Efi}$  may be negligible in SiGe HBTs, at least in those with graded Ge profiles (i.e. low barrier at the BE junction), and in the practically relevant bias region. However,  $\Delta q_{BCfi}$  is crucial for properly modeling the  $V_{CE}$  dependence of the transfer current at higher current densities. Figure 10.4 illustrates the two branches of the solution for  $i_{Tf}$  that are obtained at high-current densities. Around  $I_{CK}$  the term  $\Delta q_{BCfi}$  causes the solution to transition from  $i_{Tfi}$  of (10.31) to a lower current and temporary also lower transconductance. This was tried in [7] with the high-current correction. At very high injection  $i_{Tf}$  keeps increasing again roughly proportional to  $\exp(v_{B'E'}/(2V_T))$  due to base conductivity modulation or  $P \sim N \gg N_{BASE}$  at high injection levels and a charge saturated collector region.

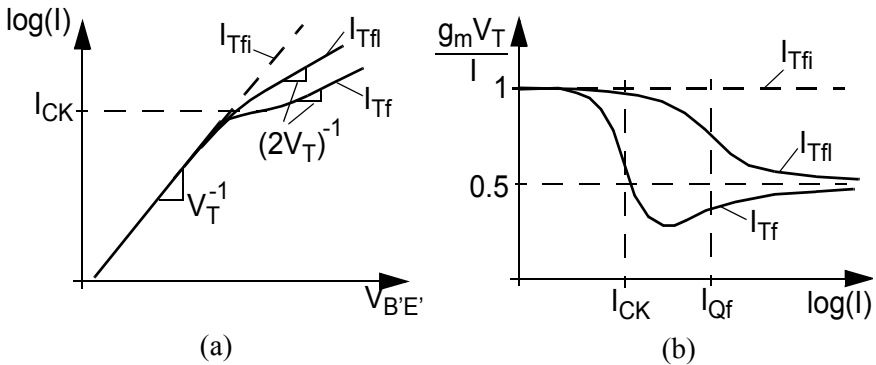


Fig. 10.4: Illustration of (a) the different transfer current branches at high injection (neglecting the emitter contribution) and (b) the corresponding normalized transconductance. In the latter, the asymptotic value 0.5 only holds without the emitter contribution.

Neglecting for the time being the emitter contribution, (10.40) reduces to

$$q_{pT}^2 - q_j q_{pT} - q_{fI} - \Delta q_{BCfi}(q_{pT}) = 0, \quad (10.43)$$

which can be solved directly for a suitable bias dependence of  $\Delta q_{BCfi}$ . The bias dependence of  $w^2$  in  $\Delta q_{BCfi}$  causes the minority charge to transition from the low-current expression  $q_{fI}$  to the high-current expression

$$q_{fh} = q_{fI} + \Delta q_{BCfi}(w = 1) = \frac{i_{Tfi}}{I_{Qf}} + \frac{i_{Tri}}{I_{Qr}} + \frac{i_{Tfi}}{I_{Qfh}}. \quad (10.44)$$

Replacing  $q_{fI}$  in (10.30) by above expression gives again at very high injection the desired dependence of  $q_{pT}$  proportional to  $I_{Tfi}^{0.5}$  as shown in Fig. 10.4. To solve the problem of a negative transconductance,  $q_{pT}$  must be prevented from increasing faster than  $I_{Tfi}$  during the transition from its low- to high-current expression. In other words, the selected smoothing function must guarantee that  $\Delta q_{BCfi}$  overall does not increase faster than  $I_{Tfi}^{m_{qm}}$  with  $m_{qm} < 2$ ; the limiting case  $m_{qm} = 2$  would result in  $g_m = 0$  somewhere within the transition range.

Evaluating  $w$  in (10.10) with  $I_{Tfi}$  instead of  $I_{Tf}$  and then solving the quadratic equation (10.43) guarantees  $q_{pT} \sim I_{Tfi}^{0.5}$  (i.e.  $g_m > 0$ ) for the asymptotic case of very high injection and also increases the allowed  $I_{Qfh}/I_{CK}$  range by at least an order of magnitude. However, in the worst case of infinite  $V_{Er}$  still a negative  $g_m$  can occur for values  $I_{Qfh}/I_{CK}$  lower than about 0.2. For realistic values of  $V_{Er}$  this ratio becomes much smaller.

In general, the goal to be accomplished is to set

$$q_f = q_{fI} + (q_{fh} - q_{fI}) s_f(I_{Tfi}, I_{CK}) \quad (10.45)$$

with  $s_f$  as smooth transition function between the minority charge branches. The function  $s_f$  must depend on  $I_{Tfi}$  (or  $V_{B'E'}$ ), the critical current  $I_{CK}$ , and has to fulfill the condition  $g_m > 0$ , which is equivalent to  $dq_{pT}/dV_{B'E'} < q_{pT}/V_T$ . Unfortunately, even for the most simple case of infinite  $V_{Er}$  the resulting expression is too complicated to be of any help for constructing an analytical expression for  $s_f$ . In [9, 10]  $s_f = w^2$  was chosen, but is evaluated in (10.10) instead of with  $i$  with the modified argument

$$i_{sf} = \frac{i_{Tfi} - I_{CK}^* q_{pT,l}}{i_{Tfi} + I_{CK}^* (q_{pT,h} - q_{pT,l})} \quad (10.46)$$

and with  $a_{hc}$  set to 0.01 in (10.10).  $q_{pT,h}$  results from replacing  $q_{fi}$  in (10.30) by  $q_{fh}$ . For providing some flexibility to adjust the transconductance and decouple the DC from the AC behavior the model parameter  $a_{hq}$  was introduced that leads to a modified critical current  $I_{CK}^* = I_{CK}(1 + a_{hq})$ .

The derivation of (10.46) in [10] is based on (10.9), (10.10) and the assumption  $I_{Tf} > I_{CK}$ . Fundamentally though,  $\Delta q_{BCfi}$  represents the charge increase from both high injection into the collector (more relevant for BJTs) and BC barrier effect (mostly relevant in HBTs). The existing function (10.10) used for  $w$  is just a mathematical means that turned out to be quite accurate for modeling the transit time. Also, the independent variable  $i$  from (10.9) results from the low-voltage solution for the collector injection width. Therefore, *any suitable* smoothing function should be usable that allows an as flexible and accurate description of high injection as possible. Furthermore, the derivation of (10.45) in [10] is somewhat confusing since  $w$  in [10, eq. (20)], which is similar to  $s_f$  in (10.46), has the meaning of a smoothing function that is limited to  $[0, 1]$ . However, it is interpreted as the normalized injection width although physically no such relation exists between  $q_{pT}$  and  $w$ . In principle, it is just coincidence that  $w$  provides the basic features needed for the purpose pursued here.

The final expression for  $q_{fh}$  is obtained in the existing model version 1.2 by adding to (10.44) the emitter term

$$\Delta q_{Efi} = \left( t_{fh} \frac{i_{Tfi}^2}{I_{CK} I_{Qfh}} \right)^{g_{Ef}} \quad (10.47)$$

with the condition  $g_{Ef} < 1$  in order to avoid  $g_m = 0$  for the asymptotic case of very high injection. In [10],  $g_{Ef} = 2/3$  is proposed, which results from assuming  $q_{pT} = \Delta q_{Efi}$  at *very high* injection and setting  $g_{IE} = 1$  initially. Due to (i) the simplifications made for the minority charge, (ii) the introduction of a smoothing function for the transition region, and (iii) the resulting limited validity of HICM/L0 for the high-current region it remains

to be clarified whether the emitter charge term is really needed for *practical* applications of this model.

The final expression for the normalized charge, including a rough description of high-current effects, can then be obtained again by solving a quadratic equation and yields:

$$q_{pT} = \frac{q_j}{2} + \sqrt{\left(\frac{q_j}{2}\right)^2 + q_{fI} + \Delta q_{BCfi}(s_f) + \Delta q_{Efi}} \quad (10.48)$$

Since above solution does not appear to generate negative  $g_m$  values for a wide and practically relevant range of parameters, it has been implemented in version 1.2. Note, that for numerical stability  $q_{pT}$  always needs to be greater than 0, which corresponds to avoiding base reach-through for negative junction voltages.

Although the existing version of HICUM/L0 includes a consistent description of the high-current region, the latter is still quite rudimentary and of limited accuracy. Deviations to measured data can partially be compensated for by the model parameters  $I_{Qfh}$ ,  $a_{hq}$ ,  $I_{Qf}$ , and  $t_{fh}$ , but this can still lead to significant deviations in the  $I_C(V_{BE})$  and  $I_C(V_{CE})$  characteristics at high current densities and toward low voltages. Therefore, if an accurate description of the transfer current characteristics, including transconductance and output conductance at high current densities and/or low  $V_{CE}$  voltages is required, it is recommended to use HICUM/L2.

### 10.2.1.5 Static base current components

The *perimeter and bottom* BE junction related components are merged into a single diode current,

$$i_{jBE} = I_{BES} \left[ \exp\left(\frac{v_{B'E}}{m_{BE} V_T}\right) - 1 \right] + I_{RES} \left[ \exp\left(\frac{v_{B'E}}{m_{RE} V_T}\right) - 1 \right]. \quad (10.49)$$

The first component represents backinjection into the emitter and the second one recombination in the BE SCR. The saturation currents follow from HICUM/L2 parameters ((8.62) along with perimeter components) as:

$$I_{BES} = I_{BEiS} + I_{BEpS} \quad \text{and} \quad I_{RES} = I_{REiS} + I_{REpS} \quad (10.50)$$

For the diode current  $I_{jBC}$  injected into the collector across the internal and external BC junction area the usual equation is employed as in (8.63),

$$i_{jBC} = I_{BCiS} \left[ \exp\left(\frac{v_{B'C}}{m_{BC} V_T}\right) - 1 \right] + I_{BCxS} \left[ \exp\left(\frac{v_{BC}}{m_{BC} V_T}\right) - 1 \right] \quad (10.51)$$

with  $I_{BCiS} = (1 - f_{ibc}) I_{BCS}$ ,  $I_{BCxS} = f_{ibc} I_{BCS}$ , and  $I_{BCS}$  as saturation current of the total BC junction. Usually, this diode current is negligible but it provides a DC path between the nodes B' and C', which sometimes aids convergence and also should be regarded as a flag for the designers if the transistor enters an undesired operating region outside the model's validity range, such as hard saturation.

### 10.2.1.6 Avalanche current

The weak avalanche current in HICUM/L2 is described by (8.65) and (8.66) with the model parameters  $f_{AVL}$  and  $q_{AVL}$  that depend on emitter area, ionization coefficients, and temperature. Since in the simplified model the zero-bias internal BC depletion capacitance  $C_{jCi0}$  may not be known it has to be eliminated from (8.66). Therefore, the new model parameter

$$e_{AVL} = \frac{q_{AVL}}{C_{jCi0} V_{DCi}} = \frac{\epsilon b_n}{2 \bar{C}_{jCi0} V_{DCi}} \quad (10.52)$$

is introduced, which lumps together various original model parameters and does not depend on emitter area. Defining also the model parameter

$$k_{AVL} = f_{AVL} V_{DCi} \quad (10.53)$$

permits us to rewrite (8.66) as

$$i_{AVL} = I_{Tf} \frac{k_{AVL}}{c_c^{1/z_{Ci}}} \exp \left( -e_{AVL} c_c^{\left(\frac{1}{z_{Ci}} - 1\right)} \right) \quad (10.54)$$

with the voltage dependent function  $c_c = C_{jCi}^* / C_{jCi0}^*$ . Above equation is continuously differentiable over bias as discussed in chapter 8. The two model parameters can always be calculated from HICUM/L2 parameters, if available.

The present formulation still depends on parameters determined from small-signal operation. A possible way for decoupling would be to set  $z_{Ci} = 0.5$  and define  $V_{DCav}$  as new model parameter that is used to adjust the voltage dependent curvature (via  $c_c$ ) of the weak avalanche current.

### 10.2.1.7 Parasitic substrate transistor

As discussed in sec. 5.4.10 a parasitic substrate transistor can be turned on depending on the transistor structure, layout, and electrical conditions. Following (5.213) the transfer current of the substrate transistor is expressed as

$$I_{TS} = I_{TSS} \left[ \exp\left(\frac{v_{BC''}}{m_S V_T}\right) - \exp\left(\frac{v_{SC''}}{m_S V_T}\right) \right]. \quad (10.55)$$

Note that the same non-ideality coefficient is being used for forward and inverse operation of the substrate transistor.

The current flowing across the total collector substrate junction is described by the usual diode equation (cf. (8.100)):

$$i_{jSC} = I_{SCS} \left[ \exp\left(\frac{v_{SC''}}{m_{SC} V_T}\right) - 1 \right]. \quad (10.56)$$

A conducting CS diode is a good indication for circuit designers that the transistor enters an undesired operating region. Also, the CS diode aids DC convergence.

### 10.2.1.8 Series resistances

The equivalent circuit of Fig. 10.2 contains the bias independent series resistances for the external collector region,  $R_{Cx}$ , and for the emitter region,  $R_E$ . In contrast to HICUM/L2, the EC contains only a single base resistance element representing the *total* base resistance

$$R_B = R_{Bi} + R_{Bx}, \quad (10.57)$$



in which  $R_{Bx}$  is a parameter and  $R_{Bi}$  is modeled separately as function of bias (see below). The physical meaning of the above series resistance components is the same as for HICUM/L2.

Starting point for the internal base resistance formulation is the description of the normalized charge given in (8.71) for HICUM/L2,

$$q_{rb} = 1 + \frac{Q_{jEi} + Q_{jCi} + Q_f + Q_r}{Q_{rb0}} \quad (10.58)$$

with  $Q_{rb0}$  ( $= Q_{p0} + \Delta Q_{rb0}$ ) as modified “zero-bias” charge due to mobility changes with bias (cf. sec. 8.1.8, [18]). Since the actual values of the internal junction charges are not separately available in HICUM/L0, the corresponding charge ratios have to be replaced by available variables. At low current densities, the internal base sheet resistance is only voltage dependent via the depletion charge ratio. Using the junction charges normalized to the respective zero-bias capacitance the charge ratio is approximated here by

$$\frac{Q_{jEi} + Q_{jCi}}{Q_{rb0}} \cong \frac{v_{jE}}{V_{r0E}} + \frac{v_{jCi}}{V_{r0C}} \quad (10.59)$$

with the functions  $v_{jE} = Q_{jE}/C_{jE0} \approx Q_{jEi}/C_{jEi0}$  and  $v_{jCi} = Q_{jCi}/C_{jCi0}$ . Furthermore,  $V_{r0E} = Q_{rb0}/C_{jEi0}$  and  $V_{r0C} = Q_{rb0}/C_{jCi0}$  are model parameters, which are defined similarly as that of the Early voltages used in the transfer current expression. Note, that the bias dependence of  $v_{jE}$  differs from that of  $v_{jEi}$ , but the impact of this difference on (10.59) is compensated for by adjusting  $V_{r0E}$  properly, e.g., to the measured bias dependence of the internal base *sheet* resistance. For homojunction transistors  $V_{r0C} = V_{Ef}$  and  $V_{r0E} = V_{Er}$  due to  $h_{jCi} = 1$  and  $h_{jEi} = 1$ .

In analogy to the simplification of the current dependent weighted charge ratios in the transfer current expression, the ratio containing the actual minority charges in (10.59) can be simplified by introducing “critical” currents  $I_{rBif}$  and  $I_{rBir}$ ,

$$\frac{Q_f + Q_r}{Q_{rb0}} \cong \frac{I_{Tf}}{I_{rBif}} + \frac{I_{Tr}}{I_{rBir}}, \quad (10.60)$$

with  $I_{rBif} = Q_{rb0}/\tau_{f0}$  and  $I_{rBir} = Q_{rb0}/\tau_r$ . This approximation does not include the impact of high-current effects. Since at high current densities HICUM/L0 becomes less accurate and the importance of  $R_{Bi}$  in general decreases, the number of model parameters is further reduced by setting  $I_{rBif} = I_{Qf}$  and  $I_{rBir} = I_{Qr}$ . The internal base resistance that results from conductivity modulation only is therefore given by (8.72) (as in HICUM/L2)

$$r_i = r_{Bi0} \frac{2}{q_{rb} + \sqrt{q_{rb}^2 + a_{qrb}}} \quad (10.61)$$

with  $a_{qrb} = 0.01$  as fixed constant, the approximated charge ratio

$$q_{rb} = 1 + \frac{v_{jE}}{V_{r0E}} + \frac{v_{jCi}}{V_{r0C}} + \frac{I_{Tf}}{I_{Qf}} + \frac{I_{Tr}}{I_{Qr}}, \quad (10.62)$$

and  $r_{Bi0}$  as a model parameter that is a function of the zero-bias sheet resistance  $r_{SBi0}$ , the emitter size, and the configuration (cf. sec. 8.1.8).

The effect of emitter current crowding can in general be described by the function (cf. sec. 8.1.8)

$$\psi(\eta) = \frac{\ln(1 + \eta)}{\eta}. \quad (10.63)$$

The current crowding factor given by (8.73),

$$\eta = f_{geo} \frac{r_i I_{Bi}}{V_T}, \quad (10.64)$$

depends on  $r_{SBi0}$ , the internal base current and emitter geometry via the factor  $f_{geo}$ . Due to the model simplifications,  $I_{Bi}$  is not directly available so that  $\eta$  has to be calculated from the base *terminal* current. This is done by modifying the geometry factor  $f_{geo}$  to

$$f_{geo}^* = \frac{f_{geo}}{1 + \gamma_B (P_{E0}/A_{E0})} \quad (10.65)$$

which now contains the base current bottom-to-perimeter partitioning (at low current densities). As a result

$$\eta = f_{geo}^* \frac{r_i I_B}{V_T} . \quad (10.66)$$

Note that  $f_{geo}^*$  does not depend on bias and, therefore, is a model parameter that can be directly calculated in advance from, e.g., HICUM/L2 parameters or measured data for a given (transistor) geometry. The final equation for the internal base resistance (cf. (8.69)) then reads

$$r_{Bi} = r_i \psi(\eta) \quad (10.67)$$

and contains the model parameters  $r_{Bi0}$ ,  $V_{r0E}$ ,  $V_{r0C}$  and  $f_{geo}^*$ . For advanced technologies with very low base sheet resistances and narrow emitter widths it is questionable though whether including emitter current crowding in a simplified model is really necessary.

Since in the simplified EC the internal base resistance cannot be accessed anymore through a separate node (between  $R_{Bx}$  and  $R_{Bi}$ ) lateral NQS effects are omitted in HICUM/L0. The corresponding NQS model is valid for small-signal operation only, and the effect usually comes into play at quite high frequencies, where HICUM/L0 becomes less accurate anyway.

### 10.2.1.9 External (parasitic) capacitances

In addition to the bias dependent junction and diffusion capacitances, advanced processes contain constant capacitances, that are caused by the isolation between base, emitter, and collector. The most relevant capacitance elements  $C_{BEpar}$  and  $C_{BCpar}$  have been added to the equivalent circuit in Fig. 10.2. Their values are usually the same as for HICUM/L2. These capacitances can also include contributions from the metallization above the silicon surface. Note that in Fig. 10.2  $C_{BCpar}$  is connected to the internal C node assuming that the major contribution to the capacitance comes from the shallow trench isolating base and collector region.

### 10.2.1.10 Self-heating

In order to make HICUM/L0 applicable for an as large as possible variety of technologies a first-order thermal network for modeling self-heating is provided as shown in Fig. 10.2. It consists of a thermal resistance  $R_{th}$

and a thermal capacitance  $C_{th}$ . The temperature dependence of the thermal conductance in  $R_{th}$  is roughly taken into account by

$$R_{th}(T) = R_{th}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{rth}} \quad (10.68)$$

with  $R_{th}(T_0)$  and  $\zeta_{rth}$  as model parameters. For modeling thermal coupling, it is recommended in a simulator implementation to make the temperature node accessible to the external circuit.

### 10.2.2 Temperature dependence

The temperature dependent modeling in HICUM/L0 follows closely the equations of HICUM/L2 (cf. section 8.3), attempting to maintain an as strong as possible relation to physics despite the simplifications. Below,  $T_0$  is again the reference temperature for which the model parameters have been determined. The formulas are expected to be valid for a temperature range between about 250K and 450K. This range depends somewhat on the technology considered. Since effects such as freeze-out are usually not taken into account by compact models, it is not recommended to use a model below about 250K unless its parameters have been extracted especially for the low-temperature range. As HICUM/L0 is a fairly new model there have been only limited results available so far on its temperature dependent modeling capability.

In order to allow simulations of devices fabricated in different materials and to make the model simulator-independent, the temperature dependent bandgap voltage (8.103) has been added to the model equations. The corresponding model parameters  $f_{1vg}$ ,  $f_{2vg}$ , and  $V_{gx}(0)$  (with  $x = \{E, B, C, S\}$ ) have the same meaning as already explained for HICUM/L2.

The temperature dependence of the transfer saturation current is dominated by the intrinsic carrier concentration and given by

$$I_S(T) = I_S(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{CT}} \exp \left[ \frac{V_{gB}}{V_T(T)} \left( \frac{T}{T_0} - 1 \right) \right] \quad (10.69)$$

with the bandgap voltage  $V_{gB} = V_{gB}(0)$  (averaged over the base region) and  $\zeta_{CT}$  as model parameters. The transfer current related parameter indicating base conductivity modulation at high injection is modeled as

$$I_{Qf}(T) = I_{Qf}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{iqf}}. \quad (10.70)$$

Similarly the current determined by collector and barrier charge storage reads

$$I_{Qfh}(T) = I_{Qfh}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{iqfh}} \quad (10.71)$$

The T dependence of the Early voltages results from various physics-based terms and has been simplified into a second-order polynomial:

$$V_{Ex}(T) = V_{Ex}(T_0)(1 + \alpha_{vex1}\Delta T + \alpha_{vex2}\Delta T^2) \quad (10.72)$$

with  $x = \{f, r\}$ ,  $\Delta T = T - T_0$ , and  $\alpha_{vex1}$ ,  $\alpha_{vex2}$  as relative TCs. The same parameters and T dependent expression are also applied to the corresponding voltages  $V_{r0E}$  and  $V_{r0C}$  used for  $R_{Bi}$ .

The BE saturation current components can be written as

$$I_{BES}(T) = I_{BES}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{BET}} \exp \left[ \frac{V_{gE}}{V_T(T)} \left( \frac{T}{T_0} - 1 \right) \right] \quad (10.73)$$

with the model parameters  $V_{gE} = V_{gE}(0)$  and  $\zeta_{BET}$ , and

$$I_{RES}(T) = I_{RES}(T_0) \left( \frac{T}{T_0} \right)^{m_g/2} \exp \left[ \frac{V_{gBE}}{V_T(T)} \left( \frac{T}{T_0} - 1 \right) \right] \quad (10.74)$$

with  $m_g = 3 - f_{avg}q/k_B$  from (6.9) and the average bandgap voltage of base and emitter,  $V_{gBE} = (V_{gB} + V_{gE})/2$ .

The temperature dependent saturation current of the BC diode reads

$$I_{BCS}(T) = I_{BCS}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{BCI}} \exp \left[ \frac{V_{gC}}{V_T(T)} \left( \frac{T}{T_0} - 1 \right) \right] \quad (10.75)$$

with the model parameter  $V_{gC} = V_{gC}(0)$  and  $\zeta_{BCI} = m_g + 1 - \zeta_{Ci}$ . The parameter  $\zeta_{Ci}$  determines the temperature dependence of  $r_{Ci0}$  in (8.112).

The temperature dependent saturation current of the SC diode reads

$$I_{SCS}(T) = I_{SCS}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{SCT}} \exp \left[ \frac{V_{gS}}{V_T(T)} \left( \frac{T}{T_0} - 1 \right) \right] \quad (10.76)$$

with  $\zeta_{SCT} = m_g - 1.5$  and the model parameter  $\mathbf{V}_{gS} = \mathbf{V}_{gS}(\mathbf{0})$ . Finally, the transfer saturation current of the substrate transistor is given by

$$I_{TSS}(T) = I_{TSS}(T_0) \left( \frac{T}{T_0} \right)^{\zeta_{SCT}} \exp \left[ \frac{V_{gC}}{V_T(T)} \left( \frac{T}{T_0} - 1 \right) \right]. \quad (10.77)$$

Employing a separate average effective bandgap voltage for each transistor region allows accurate temperature modeling.

The temperature dependence of all zero-bias junction capacitances is given by

$$C_{jx0}(T) = C_{jx0}(T_0) \left( \frac{V_{Dx}(T_0)}{V_{Dx}(T)} \right)^{z_x}. \quad (10.78)$$

with  $x = \{E, Ci, Cx, S\}$ . In addition, the parameter  $a_{jE}$  is modeled temperature dependent as

$$a_{jE}(T) = a_{jE}(T_0) \frac{V_{DE}(T)}{V_{DE}(T_0)}. \quad (10.79)$$

In order to avoid numerical issues at very high temperatures, the extended formulation (8.104) to (8.107) is used for the built-in voltages.

The temperature dependent formulations of the BC avalanche current related parameters read

$$k_{AVL}(T) = k_{AVL}(T_0) \frac{1 - \alpha_{ava} \Delta T}{1 - \alpha_{avb} \Delta T} \frac{V_{DCi}(T)}{V_{DCi}(T_0)}, \quad (10.80)$$

$$e_{AVL}(T) = e_{AVL}(T_0) [1 - \alpha_{avb} \Delta T] \left( \frac{V_{DCi}(T)}{V_{DCi}(T_0)} \right)^{z_{Ci} - 1}. \quad (10.81)$$

Finally, the temperature dependence of the parameters for the transit time and series resistances is modelled the same way as in HICUM/L2.

### 10.2.3 Small-Signal Operation

The small-signal model for HICUM/L0 is derived following the same methodologies as described in ch 8.5 for HICUM/L2. Using a Verilog-A compiler (e.g. [13, 14]) all derivatives of current and charge sources in the large-signal EC of Fig. 10.2 are generated symbolically to yield the corre-

sponding conductances, transconductances, capacitances, and trans-capacitances for a complete small-signal EC. The most important elements are displayed in the approximate small-signal EC in Fig. 10.5. The forward transconductance and output conductance are defined as

$$g_m = \left. \frac{dI_{Tf}}{dV_{B'E'}} \right|_{V_{CE}}, \quad g_o = \left. \frac{dI_{Tf}}{dV_{CE}} \right|_{V_{B'E'}}. \quad (10.82)$$

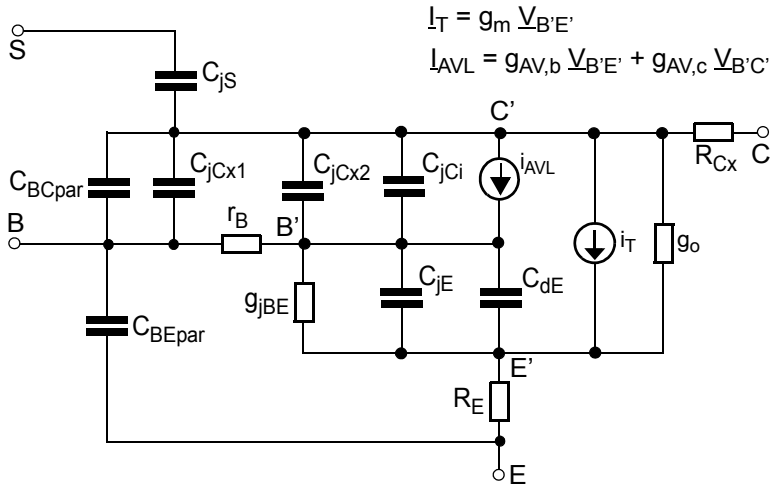


Fig. 10.5: Approximate small-signal EC of HICUM/L0 containing the most important elements for circuit design and for forward operation at  $V_{CE'} > 4 V_{CEs}$ . Elements resulting from self-heating and the parasitic substrate transistor have been omitted.

The BE junction diode related conductance is given by

$$g_{jBE} = \frac{dI_{jBE}}{dV_{B'E'}}, \quad (10.83)$$

and the avalanche current related (trans)conductances read

$$g_{AV,b} = \left. \frac{dI_{AVL}}{dV_{B'E'}} \right|_{V_{B'C}}, \quad g_{AV,c} = \left. \frac{dI_{AVL}}{dV_{B'C'}} \right|_{V_{B'E'}}. \quad (10.84)$$

The capacitances result from the derivatives of the charges, with the BE diffusion capacitance given with good accuracy by  $C_{DE} = \tau_f g_m$ . Finally, the base resistance is given by

$$r_B = R_{Bx} + r_{Bi} \quad (10.85)$$

with  $r_{Bi}$  as small-signal internal base resistance (cf. (8.140)).

#### 10.2.4 Noise Model

The noise EC of HICUM/L0 is based on its small-signal model. The correlation between base and collector current noise is neglected. The included noise sources are described below.

For series resistors, thermal noise is taken into account,

$$\overline{I_R^2} = \frac{4kT\Delta f}{R}, \quad (10.86)$$

with  $R = R_E, R_{Cx}$  or  $R_B$ .

For the transfer current, shot noise is assumed:

$$\overline{I_T^2} = 2qI_T \Delta f. \quad (10.87)$$

The noise resulting from the current injected across the BE junction into the emitter contains a shot noise and a flicker noise contribution:

$$\overline{I_{BE}^2} = 2qI_{jBE}\Delta f + k_F I_{jBE}^{a_F} \frac{\Delta f}{f}. \quad (10.88)$$

The currents across the other junctions are assumed to have a shot noise component only,

$$\overline{I_{jx}^2} = 2qI_{jx} \Delta f, \quad (10.89)$$

with  $x = \{BC, CS\}$ . Avalanche generation within the internal BC depletion region is assumed to have shot noise. Following (7.78) yields for *weak* impact ionization

$$\overline{I_{AVL}^2} = 2qMI_{AVL} \Delta f \quad (10.90)$$

with  $M = 1 + I_{AVL}/I_T$  and  $I_{AVL}$  from (10.54).



### 10.2.5 Parameter List

The following list summarizes the model parameters of HICUM/L0 v1.2. Since the model is continuously being improved to keep pace with the rapidly progressing process technology the reader is referred to the official website [6] for the most up-to-date model version.

name	description	range	unit
	transfer current		
is	Saturation current		A
mcf	Forward ideality factor	(0, 10]	-
mcr	Inverse ideality factor	(0, 10]	-
vef	Forward Early voltage	(0, $\infty$ ]	V
ver	Inverse Early voltage	(0, $\infty$ ]	V
iqf	Forward high-injection knee current (base conductivity modulation)	(0, $\infty$ ]	A
fiqf	flag for turning on voltage dependence of iqf	[0, 1]	-
iqr	Inverse high-injection knee current (base conductivity modulation)	(0, $\infty$ ]	A
iqfh	Forward high-injection knee current (collector conductivity modulation)	(0, $\infty$ ]	A
tfh	High-injection transit time ratio	[0, $\infty$ )	s
ahq	DC factor for critical current		-
	DC depletion functions for transfer current		
vdcdc	BE built-in potential		V
zedc	BE grading coefficient		-

name	description	range	unit
ajedc	Ratio of maximum to zero-bias BE capacitance	$[1, \infty)$	-
	Base current components		
ibes	BE saturation current	$[0, 1]$	A
mbe	BE current ideality factor	$(0, 10]$	-
ires	BE recombination saturation current	$[0, 1]$	A
mre	BE recombination current ideality factor	$(0, 10]$	-
ibcs	BC saturation current	$[0, 1]$	A
mbc	BC current ideality factor	$(0, 10]$	-
fibc	partitioning factor for ibcs	$[0, 1]$	-
	Base-collector avalanche current		
kavl	Avalanche current prefactor	$[0, \infty)$	
eavl	Avalanche current exponent factor	$[0, \infty)$	
	Depletion charges and capacitances		
cje0	BE zero-bias depletion capacitance	$(0, \infty)$	F
vde	BE built-in potential	$(0, 10]$	V
ze	BE grading coefficient	$(0, 1]$	-
aje	Ratio of maximum to zero-bias BE capacitance	$[1, \infty)$	-
cjci0	Internal BC zero-bias depletion capacitance	$(0, \infty)$	F
vdci	Internal BC built-in potential	$(0, 10]$	V
zci	Internal BC grading coefficient	$(0, 1]$	-

name	description	range	unit
vptci	Internal BC punch-through voltage	(0, 100]	V
cjcx0	External BC zero-bias depletion capacitance	[0, $\infty$ )	F
vdcx	External BC built-in potential	(0, 10]	V
zcx	External BC grading coefficient	(0, 1]	-
vptcx	External BC punch-through voltage	(0, 100]	V
fbc	Partitioning factor for external BC capacitance	[0, 1]	-
cjs0	CS zero-bias depletion capacitance	[0, $\infty$ )	F
vds	CS built-in potential	(0, 10]	V
zs	CS grading coefficient	(0, 1]	-
vpts	CS punch-through voltage	(0, 100]	V
	Minority charge storage effects		
t0	Low-current forward transit time at $V_{BC}=0V$	[0, $\infty$ )	s
dt0h	Time constant for base and BC space charge layer width modulation		s
tbvl	Time constant for modelling carrier jam at low VCE	[0, $\infty$ )	s
tef0	Neutral emitter storage time	[0, $\infty$ )	s
gte	Exponent factor for current dependence of neutral emitter storage time	[0, 10]	-
thcs	Saturation time constant at high injection	[0, $\infty$ )	s
ahc	Smoothing factor for current dependent of base and collector transit time	(0, 10]	-

name	description	range	unit
rci0	Low-field internal collector resistance	$(0, \infty)$	$\Omega$
vlim	Voltage separating ohmic and saturation velocity regime	$(0, 10]$	V
vc es	Internal CE saturation voltage	$[0, 1]$	V
vpt	Collector punch-through voltage	$(0, 100]$	V
tr	Storage time for inverse operation	$[0, \infty)$	s
	Series resistances		
rb i0	Zero-bias internal base resistance	$[0, \infty)$	$\Omega$
vr0e	Forward Early voltage	$(0, \infty]$	V
vr0c	Inverse Early voltage	$(0, \infty]$	V
fgeo	Geometry factor for E current crowding	$[0, \infty]$	-
rbx	External base series resistance	$[0, \infty)$	$\Omega$
re	Emitter series resistance	$[0, \infty)$	$\Omega$
rcx	External collector series resistance	$[0, \infty)$	$\Omega$
	Substrate transistor		
itss	Saturation current of substrate transistor transfer current	$[0, 1]$	A
msf	Ideality factor of substrate transfer current	$(0, 10]$	-
iscs	Saturation current of CS diode	$[0, 1]$	A
m sc	Ideality factor of CS diode	$(0, 10]$	-
	Parasitic isolation capacitances		
cbepar	Total parasitic BE capacitance	$[0, \infty)$	F

name	description	range	unit
cbcpar	Total parasitic BC capacitance	$[0, \infty)$	F
	Noise		
kf	Flicker noise coefficient (no unit only for AF=2)	$[0, \infty)$	-
af	Flicker noise exponent factor	$(0, 10]$	-
	Temperature dependence		
vgb	Effective base region bandgap-voltage	$(0, 10]$	V
vge	Effective emitter bandgap voltage $V_{gEff}$	$(0, 10]$	V
vgc	Eff. collector bandgap voltage $V_{gCeff}$	$(0, 10]$	V
vgs	Eff. substrate bandgap voltage $V_{gSeff}$	$(0, 10]$	V
f1vg	Coefficient $K_1$ in bandgap equation		V/K
f2vg	Coefficient $K_2$ in bandgap equation		V/K
zetact	Exponent coefficient in transfer current temperature dependence	$[-10:10]$	-
zetabet	Exponent coefficient in BE junction current temperature dependence	$[-10:10]$	-
alt0	First-order relative TC of parameter $t_0$	$[-10:10]$	1/K
kt0	Second-order relative TC of parameter $t_0$	$[-10:10]$	$1/K^2$
zetaci	Temperature exponent for $r_{Ci0}$	$[-10:10]$	-
alvs	Relative TC of saturation drift velocity	$[-10:10]$	1/K
alces	Relative temperature coefficient of $V_{CEs}$	$[-10:10]$	1/K

name	description	range	unit
zetarbi	Temperature exponent of internal base resistance	[-10:10]	-
zetarbx	Temperature exponent of external base resistance	[-10:10]	-
zetarcx	Temperature exponent of external collector resistance	[-10:10]	-
zetare	Temperature exponent of emitter resistance	[-10:10]	-
zetaiqf	Exponent factor for iqf	[-2:2]	-
zetaiqfh	Exponent factor for iqfh	[-2:2]	-
zetarth	Exponent factor for rth	[0:5]	
alkav	Relative temperature coefficient for kav1	[-10:10]	1/K
aleav	Relative temperature coefficient for eav1	[-10:10]	1/K
	Self-Heating		
rth	Thermal resistance	[0:∞)	K/W
cth	Thermal capacitance	[0:∞)	Ws/K
flsh	flag for turning on (1)/off (0) self-heating	[0, 1, 2]	-
	circuit simulator specific parameters		
tnom	parameter temperature specification ( $T_0$ )		°C

name	description	range	unit
dt	change w.r.t. chip (substrate) temperature		°C
version	model version identifier	-	-

Table 10.1: List of HICUM/L0 model parameters and flags. Note, that all flags, the band-gap voltage parameters (f1vg, f2vg), and the simulator parameters in the last block are not counted as model parameters. The parameters TNOM and DT are available in most simulators and are also mostly named the same way. The “model version identifier” enables version control in simulators with different HICUM generations.

### 10.2.6 Parameter Extraction

Since HICUM/L0 was derived from the more sophisticated and accurate HICUM/L2, the parameter extraction can be performed in two different ways: (i) calculation of HICUM/L0 parameters directly from HICUM/L2 parameters *and* characteristics, *or* (ii) extraction from measured characteristics of, e.g., a single transistor. Below, these options are briefly discussed.

#### 10.2.6.1 Parameter determination from HICUM/L2

Many parameters can be calculated directly from the HICUM/L2 parameter set, namely those for: CS, BE, BC diode current and depletion charge, series resistances including most of the parameters for the internal base resistance, self-heating, temperature dependence (many of the coefficients), parasitic substrate transistor, transit time (most parameters), avalanche current. Also, useful initial values for the remaining parameters of the transfer current and transit time can be generated from HICUM/L2. The final values for these remaining parameters are then extracted on HICUM/L2 characteristics. Such a procedure can be automated and has been implemented in the TRADICA program [4, 11, 12].

### 10.2.6.2 Parameter extraction from experimental data

The similarity of HICUM/L0 to the SGPM at low injection allows to apply already existing extraction procedures for the latter after some extensions or modifications. Also, the simplicity of HICUM/L0 allows extractions on single (discrete) transistors. Table 10.2 contains the possible flow of a step-by-step extraction procedure that is based on [8] but has been modified and extended towards the version 1.2 described here. Due to the strong self-heating in advanced process technologies it is recommended to determine the temperature dependent parameters early in the procedure to enable possible corrections during the extraction of the other parameters. Of course, an isothermal measurement capability would alleviate this problem.

Step	EC element	Parameters
1	Total BE, BC, and CS depletion capacitances	$C_{jE0}, V_{DE}, z_E$ $C_{jCi0}, V_{DCi}, z_{Ci}, V_{PTCi}$ $C_{jCx0}, V_{DCx}, z_{Cx}, V_{PTCx}$ $C_{jS0}, V_{DS}, z_S$
2	Current components related to BE, BC, and CS junction	$I_{BES}, m_{BE}, I_{RES}, m_{RE}$ $I_{BCS}, m_{BC}$ $I_{SCS}, m_{SC}$
3	Temperature coefficients, self-heating	$\zeta_{ci}, \zeta_{rbi}, \zeta_{rbx}, \zeta_{rcx}, \zeta_{re}$ $R_{TH}, C_{TH}$
4	Series resistances	$r_{Cx}, r_E, r_{Bx}$ $r_{Bi0}, V_{r0E}, V_{r0C}, f_{geo}$
5	Transit time at low injection	$\tau_0, \tau_{Bvl}, \Delta\tau_{0h}, a_{jE}$
6	Transfer current at low and medium injection	$I_S, m_{Cf}, m_{Cr}, V_{Ef}, V_{Er}$ $V_{DEdc}, z_{Edc}, a_{jEdc}$



Step	EC element	Parameters
7	BC Avalanche current	$e_{AVL}, k_{AVL}$
8	Critical current	$r_{Ci0}, V_{CES}, V_{PT}, V_{lim}$
9	Transit time at high injection	$\tau_{Ef0}, g_{\tau E}, \tau_{hcs}, a_{hc}$
10	Transfer current at high injection	$I_{Qf}, I_{Qr}, I_{Qfh}, t_{fh}, a_{hq}$
11	Remaining parameters and fine tuning	$K_F, A_F, V_{Gb}, V_{Gc}, V_{Ge}, V_{Gs}, \zeta_{iqf}, \zeta_{iqfh}, \zeta_{rth}, \alpha_{kav}, \alpha_{eav}, \alpha_{vs}, \alpha_{ces}, \alpha_{t0}, k_{t0}$

Table 10.2: HICUM/L0 parameter extraction steps.

### 10.3 HICUM/L4: a distributed model

As shown in [3] the 3D pinch-in effect caused by BC avalanche breakdown can only be captured by a distributed model for the internal transistor region. A completely different problem is self-heating and thermal coupling (e.g., [2, 19]), for instance in power amplifiers, which tend to require large multi-finger or even multi-cell transistor structures. Other problems, which cannot be accurately described by a lumped internal base resistance or impedance, are the impact of transient emitter current crowding on large-signal switching (and probably also cyclostationary noise) behavior and of DC emitter current crowding on high-frequency *noise* behaviour. In general, relatively large emitter widths, large transistor structures or operation in the breakdown regime cause distributed thermal and electrical effects that cannot be captured by compact model forms such as HICUM/L0 and HICUM/L2 as well as any other model with similar topology and *lumped* formulations. Therefore, the transistor structure has to be discretized by applying lumped models for sufficiently small regions. This leads to a distributed model as shown in Fig. 10.1 for example. The approach of building such a model, named HICUM/L4, is briefly described below.

Offering HICUM/L4 extends the existing HICUM hierarchy and addresses the following issues:

- Final verification of critical functions in circuits by using a distributed model.
- L4 parameters can be generated from the geometry and process specific HICUM/L2 parameter set in TRADICA, for instance. Hence, in case of integrated circuit design no extra effort is required for both parameter extraction and parameter generation.
- The model equations are physics-based, enabling predictive and statistical design.

The disadvantages of HICUM/L4 are obviously increased computational effort during simulation. However, in large circuits generally only very *few critical* transistors exist that need to be described by such a sophisticated and accurate distributed model. This section contains the *basic ideas* of HICUM/L4. A detailed description would require us to list the complete set of geometry scaling equations for all possible configurations, which is beyond the scope here.

The discretization of a distributed model depends on the application. Emitter current crowding or the pinch-in effects require the discretization of the emitter (junction) *area*, i.e. of an emitter *finger*. This is illustrated in Fig. 10.6 for an arbitrary non-equidistant discretization. Each rectangle in the top view of an emitter on the left is represented by a transistor element in the EC on the right. These transistor elements are modeled by either HICUM/L2 or HICUM/L0 depending on the operating region that needs to be accurately covered. The discrete internal transistor models  $T_{nm}$  are connected via elements of the internal base resistance:

$$R_{b, mn} = \frac{r_{Sbi} \Delta b_{E, m}}{\alpha_b \Delta l_{E, n}} \quad \text{and} \quad R_{l, mn} = \frac{r_{Sbi} \Delta l_{E, n}}{\alpha_l \Delta b_{E, m}} \quad (10.91)$$

with  $\Delta b_{E, m} = (b_{E, m} + b_{E, m-1})/2$  and  $\Delta l_{E, n} = (l_{E, n} + l_{E, n-1})/2$ . The factor  $\alpha$  depends on the discretization (odd, even) and the symmetry of the device structure. Assuming for instance a double-base contact structure, only one quarter needs to be represented in the EC, so that the element values of  $T_{nm}$  represent the four corresponding rectangles and, therefore,  $\alpha = 4$ . Since the  $T_{nm}$  are placed at the center of a rectangle, the first resistor element at

the emitter perimeter edge always represents only one half of the path length.

The connection to the external transistor EC assumes an equipotential line at the emitter perimeter edge. As shown in ch. 5 for the base resistance considerations this assumption is fulfilled in symmetric contacting schemes and still appears to be quite reasonable for other schemes. Only in such cases a lumped external EC can be used. On the other hand, if significant current crowding is expected in a critical device, always a symmetric layout should be chosen in order to minimize distributed effects.

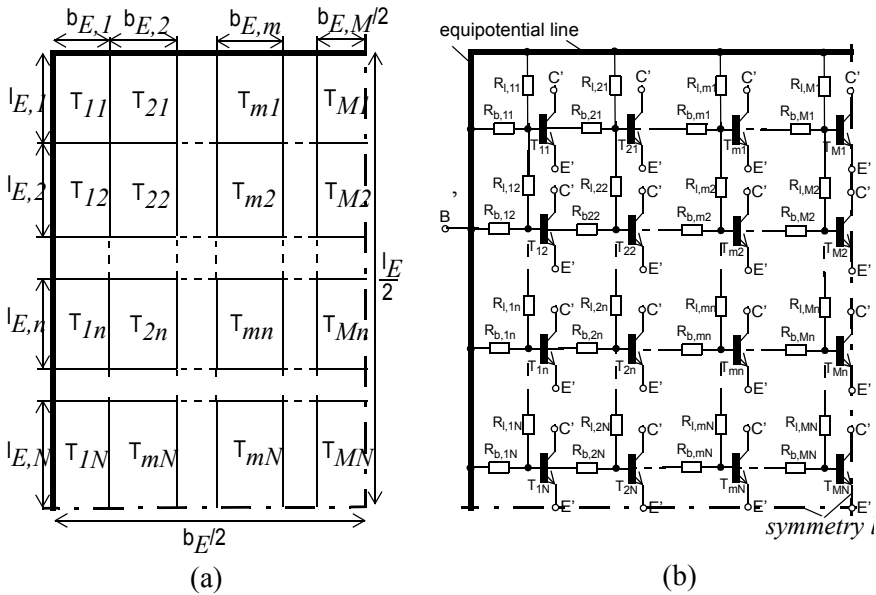


Fig. 10.6: Discretization of an emitter finger of a double-base contact structure with  $M$  ( $N$ ) elements in  $b$  ( $l$ ) direction: (a) emitter top view with discrete spatial elements; (b) HICUM/L4 EC for the internal transistor using HICUM/L0 elements.

A high discretization enables to investigate in detail the spatial distribution of electrical variables under the emitter. This is often of interest for device modeling. However, in order to minimize the computational effort during circuit design though an optimum discretization is sought that maintains sufficiently high accuracy at a minimum number of elements and nodes. Non-equidistant spacing becomes increasingly important as the

number of elements decreases. This was shown in [3] for modeling the pinch-in effect from avalanche breakdown: a non-equidistant  $b_E$  discretization (0.4:0.6) and equidistant  $l_E$  discretization (3 elements) resulted in a reduction to a 6-transistor model that still yielded acceptable accuracy. The same width distribution without any length discretization can be used for fast large-signal transients outside the high-current region [20].

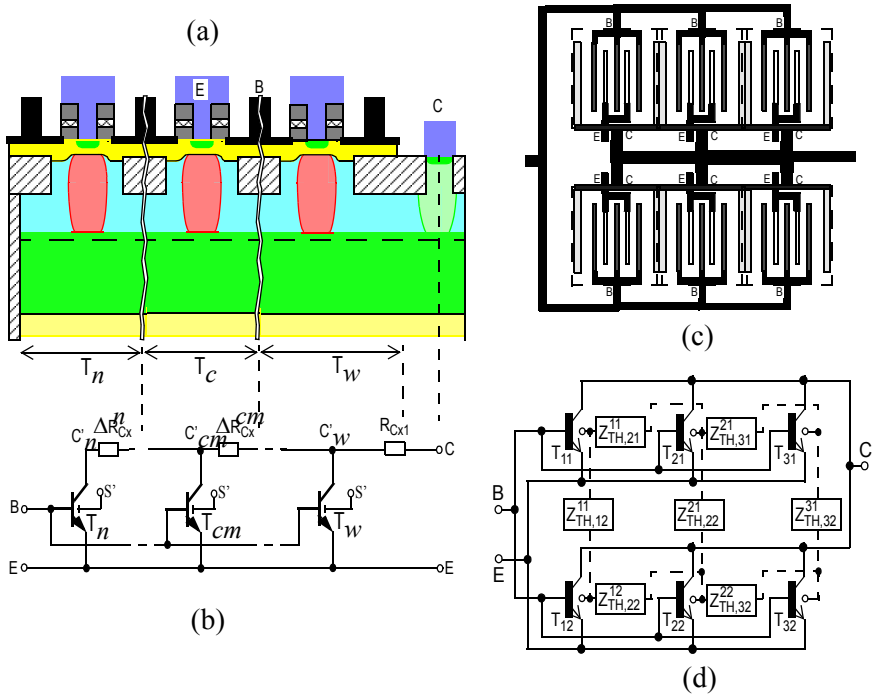


Fig. 10.7: HICUM/L4 EC with device and cell discretization: (a) cross-section through a multi-finger transistor and (b) corresponding distributed EC with one transistor element per finger. (c) Layout of power array and (d) associated distributed EC with one transistor element per basic cell in the array and self-heating coupling networks (represented by  $Z_{TH}$ ) between nearest neighbors.

In the practical implementation of the distributed model the internal base node of each transistor element should be externally accessible. While this is possible in [21], commercial circuit simulators usually do not make this node available. As a consequence, an average value (over bias) has to be inserted for  $r_{SBI}$  in (10.91) already during model parameter generation.

Electro-thermal effects occur on a large scale. Here either each emitter finger or the basic cell itself represents the smallest lumped structure for a discretization. Examples for these cases are shown in Fig. 10.7. As discussed already in ch. 5, bipolar transistors allow for quite some variation in arranging the fingers and contacts. A common multi-finger structure with a single collector contact is shown in Fig. 10.7a. Assuming long parallel fingers, fundamentally, there are certain basic elements that need to be modeled differently in terms of their geometry dependence:

- An outer element  $T_w$  containing the collector contact region.
- An outer element  $T_n$  that does not contain a collector contact region but needs to include the respective base boundary of the transistor.

Center or inner elements  $T_c$  that do not contain any elements representing the outer lateral boundaries in emitter width direction.

Figure 10.7b shows the connection of the lumped transistor element types ( $T_n$ ,  $T_c$ ,  $T_w$ ) for a structure with  $m = 1 \dots M$  inner fingers. The EC of each lumped transistor element corresponds to that of L0 in Fig. 10.2 without  $R_{Cx}$  in order to be able to access the internal collector node. For a single collector contact ( $n_C = 1$ ) each transistor element in the EC of Fig. 10.7b represents one emitter finger. For  $n_C = 2$  and an *even* number of emitter fingers each transistor element represents two fingers due to symmetry reasons. For  $n_C = 2$  and an *odd* number of emitter fingers a *single* transistor element needs to represent the center finger, while all other transistor elements again represent two fingers. Such case distinctions apply also to  $R_{Cx}$ , the parasitic capacitances, and the substrate junction related elements. Note, that the network can only be reduced due to symmetry reasons, if the selected transistor cell does not experience any thermal coupling from adjacent cells that are *not symmetrically located*. In the latter case, each stripe needs to be represented separately and thermally coupled to other stripe elements. Thermal coupling networks were already discussed in ch. 6.

Figure 10.7c shows a typical PA structure consisting of an array of basic cells, such as a 2-finger transistor. In this case, each basic cell is represented by its own model  $T_{nm}$  in the distributed electrical EC as illustrated in Fig. 10.7d. The EC of the model corresponds to that of a full HICUM/L0 or L2. The basic cells are thermally coupled through the thermal subcircuits discussed in ch. 6. In addition, such a distributed model also allows

to improve the dynamic electrical representation of the distributed structure if the metal connections are represented by transmission lines.

#### 10.4 References

- [1] M. Schröter, "High frequency circuit design oriented compact bipolar transistor modelling with HICUM", *IEICE Trans. Electron.*, Vol. E88-C, No. 6, June 2005.
- [2] D.J. Walkey, T. Smy, R. Dickson, J. Brodsky, D. Zweidinger, R. Fox, "Equivalent circuit modeling of static substrate thermal coupling using VCVS representation", *IEEE J. Solid-State Circuits*, Vol. 37, pp. 1198-1206, 2002.
- [3] M. Rickelt, H.-M. Rein, "A novel transistor model for simulating avalanche-breakdown effects in Si bipolar circuits", *IEEE J. Solid-State Circuits*, Vol. 37, pp. 1184-1197, 2002.
- [4] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener and A. Koldehoff, "Physics- and process-based bipolar transistor modeling for integrated circuit design", *IEEE Journal of Solid-State Circuits*, Vol. 34, pp. 1136-1149, 1999.
- [5] M. Schröter, S. Lehmann, H. Jiang, and S. Komarow, "HICUM/Level0 – A simplified compact bipolar transistor model," *Proc. IEEE BCTM*, pp. 112–115, 2002.
- [6] A. Mukherjee, S. Lehmann and M. Schröter, "HICUM/Level0 User's Manual", [Online]. Available at <http://www.iee.et.tu-dresden.de/iee/eb>.
- [7] M. Schröter, S. Lehmann, S. Fregonese, and T. Zimmer, "A computationally efficient physics-based compact bipolar transistor model for circuit design - Part I: Model formulation," *IEEE Trans. Electron Devices*, vol. 53, no. 2, pp. 279–286, Feb. 2006.
- [8] S. Fregonese, S. Lehmann, T. Zimmer, M. Schröter, D. Celi, B. Ardouin, H. Beckrich, P. Brenner, W. Kraus, "A Computationally Efficient Physics-Based Compact Bipolar Transistor Model for Circuit Design—Part II: Parameter Extraction and Experimental Results", *IEEE Trans. Electron Dev.*, Vol. 53, pp. 287- 295, 2006.
- [9] C. Thiele, J. Berkner, W. Klein, "Improvement of the transfer current modeling in HICUM/L0", *Proc. IEEE BCTM*, pp. 62–65, 2007.

- [10] C. Thiele, W. Klein, "A new explicit expression for the normalized hole charge improves the accuracy of HICUM/L0", *Proc. IEEE BCTM*, pp. 77–80, 2008.
- [11] M. Schröter, Y. Zimmermann, "TRADICA: Physics- and Process-Based Model Generation For Integrated Devices", *Manual Version 5.2*, 2006.
- [12] N. Derrier, "From HICUM level 2 to HICUM level 0", 8th HICUM Workshop, Böblingen, Germany, May 20-21, 2008.
- [13] [Online]. Available: [www.tiburion-da.com](http://www.tiburion-da.com)
- [14] L. Lemaitre, C. McAndrew, and S. Hamm, "ADMS – Automatic device model synthesizer," in *Proc. Custom Integrated Circuits Conf.*, 2002, pp. 27–30.
- [15] Z. Huszka, "Removing ambiguity from Hicup/L2 by an improved GICCR", 8th Hicup Workshop, Böblingen, Germany, May 20-21, 2008.
- [16] C. McAndrew, J.A. Seitchik, D.F. Bowers, M. Dunn, M. Foisy, I. Getreu, M. McSwain, S. Moinian, J. Parker, D.J. Roulston, M. Schröter, P. v. Wijnen and L.F. Wagner, "VBIC95, the vertical bipolar intercompany model", *IEEE J. Solid-State Circuits*, Vol. 31, pp. 1476-1483, 1996.
- [17] P. Borwein and T. Erdélyi, "Polynomials and Polynomial Inequalities", New York: Springer-Verlag, 1995.
- [18] H.-M. Rein and M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors at high current densities - Part II: Two-dimensional model and experimental results", *IEEE Trans. Electron Dev.*, Vol. 34, pp. 1752-1761, 1987.
- [19] Y. Zimmermann, "Modeling of spatially distributed and sizing effects in high-performance bipolar transistors", MSEE/Diploma thesis, Chair for Electron Devices and Integrated Circuits, TU Dresden, 2004.
- [20] H.-M. Rein, "Improving the Large Signal Models of Bipolar Transistors by Dividing the Intrinsic base into two Lateral Sections", *Electronics Lett.*, Vol. 13, pp. 40-41, 1977.
- [21] M. Schröter, "Transient and small-signal high-frequency simulation of numerical device models embedded in an external circuit", *COMPEL*, Vol. 10, No. 4, pp. 377-378, 1991.

This page is intentionally left blank



## **Chapter 11**

### **Application Examples**

HICUM has been used for a number of years for production and exploratory circuit design. This chapter provides an overview on the various applications of the model that is by no means complete. It starts with practical considerations of geometry scaling. Then, the achievable model accuracy is demonstrated for selected device characteristics covering DC, HF, noise and distortion characteristics. Next, a statistical and predictive modeling and simulation approach is discussed using the physics-based and geometry scaling features of the model. Finally, results for selected circuits are provided, which have been designed using HICUM and have been fabricated by various process technologies.

## 11.1 Geometry scaling approach

Geometry scalable models are important for circuit optimization through device sizing (e.g. [1, 2]), for statistical modeling and simulation (cf. section 11.2.4) as well as for simulating device matching. As mentioned in ch. 8, most of the geometry scaling related calculations can be performed as preprocessing before running the actual circuit simulation. Historically and in contrast to MOSFET models, bipolar transistor models in circuit simulators have never really included suitable geometry scaling. In other words, it has always been expected that a library is provided with model parameter values for each discrete device structure (“discrete model cards”). Presumably, the lack of a scaling preprocessor<sup>1</sup> directly associated with a compact model resulted from the difficulty of being able to develop a generally applicable accurate geometry dependent description for the large possible variety of BJT and HBT configurations (compared to MOSFETs<sup>2</sup>). It is interesting to note that the CMC [3] during the long selection process for standard BJT/HBT models (that lasted from 1999 to 2003) explicitly excluded geometry scaling as evaluation criterion. The discussion below uses the terminology and definitions associated with (2.15).

Assuming the availability of a suitable set of geometry scaling equations for a *given* process, a preprocessor then would calculate the actual model parameters for any desired transistor configuration  $c$  from specific electrical parameters  $e$  and dimensions  $d$ . The latter two vectors represent *process-independent* parameters that result from parameter extraction and design rules. Examples for possible basic transistor configurations that usually need to be covered for a given process by the scaling equations are shown in Fig. 11.1. Allowing also a variety of *predefined* substrate contact locations, obviously a quite sophisticated set of scaling equations is required to cover all possible structures.

- 
1. Many simulators offer some geometry scaling, which is extremely rudimentary though and quite inaccurate for most applications. All simulators contain a multiplication factor for modeling the replication of identical transistor cells.
  2. This statement only applies to conventional MOSFET structures for digital applications, but neither to structures for HF applications nor resulting from extreme shrinking.

If these equations are implemented in circuit simulators as *part* of a compact model they would need to be applicable for all possible processes and device configurations in order to make the model generally applicable. Considering the large possible variety of BJT/HBT structures this not only is a very difficult task but would also result in a constant code change activity to adapt the scaling equations to new configurations or process development. The associated model code maintenance effort is extremely high. An even more important issue though is that many companies view the specific data  $e$ ,  $t$ , and  $\sigma$  (cf. (2.15)) of their process as intellectual property and have little interest to release this information to their customers and competitors. This also applies to the compact models itself (or at least the scaling equations) since having better models (and design kits) is increasingly seen by foundries as competitive advantage for selling their processes.

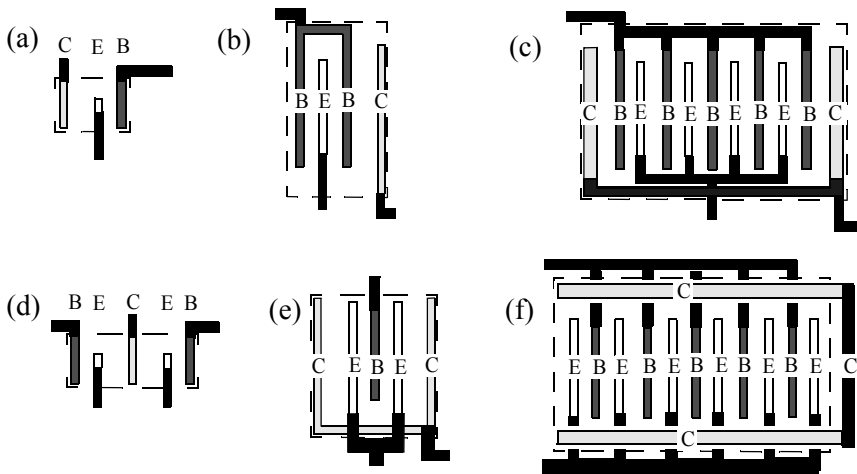


Fig. 11.1: Examples for possible transistor configurations fabricated in a given process: (a) minimum size device; (b) standard BEBC device layout; (c) standard multi-stripe double-collector layout; (d) butted-collector layout; (e) double-emitter with single base layout for PAs; (f) extension of (e) to a multi-stripe structure with perpendicular collector stripes. Note, that the location of substrate contacts is not shown here.

There are further serious disadvantages of having scaling equations associated with each compact model in a simulator:

- If different compact models (e.g. a model hierarchy) are offered for the same process the user expects consistent electrical characteristics in bias and temperature regions that should be covered by *both* models equally well. For instance, at low injection the  $I_C(V_{BE})$  or  $C_{jE}(V_{B'E'})$  curves of a SGPM should not differ from those of HICUM. However, if each model (with different model development goals and schedules) carries its own scaling equations, differences are unavoidable.
- Depending on their process capabilities, foundries often impose restrictions on transistor configurations. For instance, some foundries do not allow variable emitter widths, while others may not be able to describe their process with the given equations for perpendicular collector or base contact stripes. Realizing such restrictions would be almost impossible with model-associated scaling code but is very easy when generating the foundry-specific library.
- If a set of scaling equations would exist that is generally applicable to all models and processes one may be inclined to use Verilog-A for implementation in order to minimize model maintenance in simulators. This is only possible though for sufficiently simple equations. For instance, as soon as solutions of nonlinear equations, of (simple) differential equations, of systems of equations, or evaluations using Green's functions are required, language limitations of Verilog-A are encountered and standard programming languages such as Fortran or C are much more suitable.
- Scaling code within a circuit simulator also makes it difficult to generate distributed models since this requires a modification of the already existing netlist. Again, this problem and offering a model hierarchy is easily solved with a library.

Due to these many issues it is generally preferable to develop the scaling preprocessor separately from the circuit simulator. This way, a variety of other features can be added that are useful for not only modeling but also circuit design and process development. Such an approach has been taken with the program TRADICA [4] that started out in the early eighties as a transistor sizing program for circuit optimization [5, 1]. For accomplishing this task a set of scaling equations for a BJT process of that time was developed and implemented, and SGPM model libraries were generated. Since then, the program has been extended into many directions that are partially illustrated in Fig. 11.2:

- The number of BJT/HBT models has been extended by HICUM/L2 and, more recently, HICUM/L0 and L4. From HICUM/L2 as the reference model, consistent parameter sets for the other models (incl. the SGPM) can be generated using internal parameter extraction procedures.
- The presently existing variety of configurations comprises more than those shown in Fig. 11.1, including substrate contact locations and different isolation schemes (junction and deep trench).
- Besides a self-aligned BJT process technology several other BJT/HBT technologies, such as a selectively grown epitaxial SiGe and a mesa III/V HBT process, are described by specific scaling equations allowing a broad application of the program across foundries.

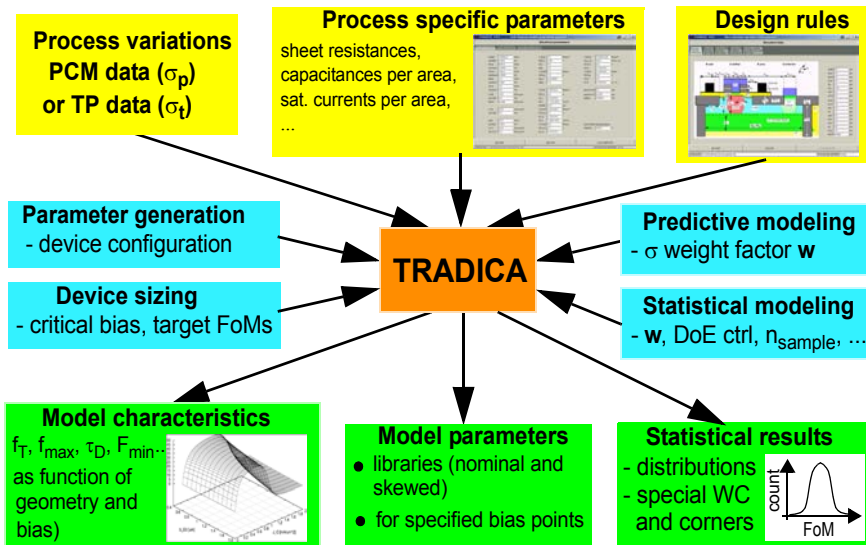


Fig. 11.2: Overview on the model parameter generation and device sizing system TRADICA (TRANsistor DIMensioning and CALculation program).

- Parameters for statistical modeling can be generated based on either process control monitor (PCM) or technology parameter (TP) data (cf. section 11.2.4). An extension of this approach and the corresponding equations facilitates the generation of model parameters for predicting the impact of larger process changes on device characteristics during process development. A minimization algorithm such as simulated

annealing allows multivariate process optimization towards device and circuit related figures of merit (FoMs).

- Different analytical device sizing algorithms can be selected depending on desired goals specified as target FoMs or critical bias points. The same multivariate optimization methods are used as for the predictive model generation approach mentioned before.
- More recently, models and scaling equations for other devices typically offered for a BiCMOS process have been implemented. Such devices include MOSFETs (described by the EKV3 model), integrated inductors, MIM capacitors and HF resistors. This also allows the users to include inter-device correlation during statistical simulation. Since the evaluation of nonlinear models in the program is not based on a Newton-algorithm, no derivatives are required thus facilitating a quick model integration and evaluation.
- Model parameter extraction has been added in the form of both generic blocks (for  $C(V)$ , diode  $I(V)$  data) and HICUM-specific procedures. The control of these procedures by simple scripts becomes less time consuming for an established extraction flow and facilitates automated extraction, especially by users that are not too familiar with HICUM.
- Managing the many features described above and the corresponding input data is accomplished by a graphical user interface (GUI). Examples for these are included in the upper part of Fig. 11.2. The GUI has been combined with that for a DD/HD device simulator [6] and a Boltzmann solver [7]. This enables the users to build an efficient and user friendly system for supporting modeling, circuit design and process development, which is independent of EDA tools and their expensive licenses.

In summary, the sophisticated set of geometry scaling equations enables a number of features that are not being offered in circuit simulators. These features are also very difficult to realize there, since EDA vendors generally have little or no expertise in device modeling. While the modeling system can control circuit simulators its features are often also desirable *within* other systems, such as scalable parameter extraction software [8] or process design kits (PDKs). Integration into the latter allows circuit designers for instance to select continuously scalable compact models directly by specifying the device configuration, to use automated circuit optimization through device sizing, and to perform statistical simulation. Note that the time for generating the model parameter sets for all the dif-

ferent device configurations in a circuit is orders of magnitude smaller than the circuit simulation itself, so that there is basically no impact in overall run time.

The geometry scaling features of HICUM have been exploited in [9-13], in addition to the TRADICA-based scalable library generation at Atmel [10, 14], Jazz [11], and users of the XMOD parameter extraction toolkit [8]. The TRADICA approach has also been applied to III/V HBTs in [15, 16, 17].

## 11.2 Examples of model comparisons with experimental data

### 11.2.1 HICUM/L2

Model parameters have been extracted for many BJT and HBT process technologies. Extensive and consistent sets of model comparisons with measured data are available in doctoral theses [18, 19, 20], foundry design manuals, which are proprietary documents though [21], in publications, and have also been regularly presented at HICUM Workshops [22]. There is certainly not enough space here to repeat even a portion of all these results, so only a few were selected with the purpose to illustrate the level of accuracy that can be achieved with HICUM. The subsequent figures and their discussion are organized according to a specific type of characteristics and model verification.

Figure 11.3 shows the collector current and transit frequency of transistors with different emitter length and at constant emitter width fabricated in a 0.18 $\mu\text{m}$  SiGe HBT BiCMOS process [19]. Excellent agreement is obtained over a wide length range thanks to a thorough parameter extraction procedure. Similar results were published in [23] for an improved process generation and in [13] for a 200 GHz 0.13 $\mu\text{m}$  SiGe BiCMOS process, both targeting optical and millimeter wave applications. Of particular interest in such advanced processes is the impact of self-heating (SH), which is shown in Fig. 11.4. According to Fig. 11.4a, voltage controlled output characteristics are affected significantly by SH as can be clearly seen by comparing the dotted lines, obtained without taking into account SH in HICUM, and the solid lines that were obtained with SH turned on.

Figure 11.4b provides a feel for the typical temperature increase in such transistors, which appears to have a limited impact on  $f_T$  though. Nevertheless, the model describes the temperature dependent changes of the characteristics quite well.

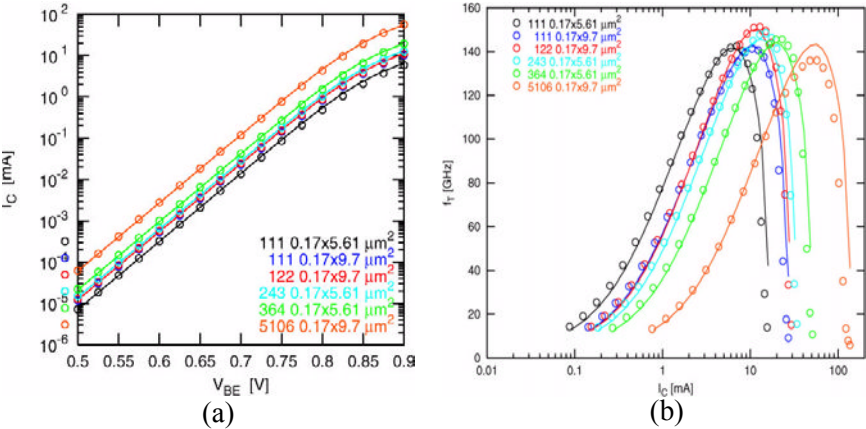


Fig. 11.3: (a) Collector current vs. BE voltage and (b) transit frequency vs. collector current at  $V_{BC} = 0$  for transistors with variable emitter length  $l_{E0}$  of a  $0.18\mu\text{m}$  SiGe HBT BiCMOS process from ST. Comparison between measurements (symbols) and HICUM (solid lines) [19].

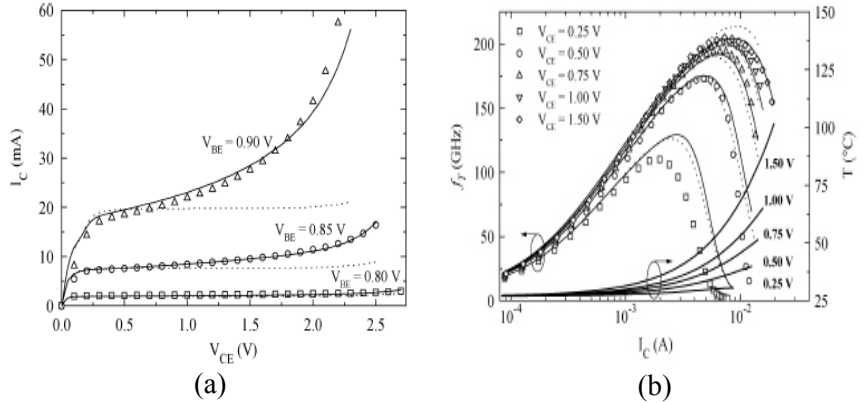


Fig. 11.4: (a) Output characteristics at constant  $V_{BE}$  and (b) transit frequency at constant  $V_{CE}$  for a  $0.17 \times 14.8\mu\text{m}^2$  SiGe HBT of a  $0.18\mu\text{m}$  BiCMOS process from ST [23]: comparison between measurements (symbols), HICUM with self-heating turned on (solid lines) and turned off (dotted lines). (© 2005 IEEE)



Besides standard DC and small-signal characteristics verification of large-signal modeling is of high interest for most applications. Since the high speed of modern HBTs makes it very difficult to directly measure the switching behavior in time domain, as it was done for older technologies [24], distortion measurement in frequency domain is a convenient way to evaluate the large-signal nonlinear behavior. Excellent agreement was obtained for BJT production technologies for harmonic distortion in a  $50\Omega$  system [25]. Similar results have been obtained for SiGe HBTs as shown in Fig. 11.5 for a 40 GHz Atmel process [26] and in Fig. 11.6 for a 100GHz IBM process [27]. In both cases though the power transistor version was used, the  $f_T$  of which is much lower than that of the corresponding high-speed version. It can be observed that both the location and amplitude of the maxima and minima in Fig. 11.5a are well captured by the model. The same holds true for power gain and IMD3 in Fig. 11.5b. For PAs both quasi-saturation and avalanche region are important to maximize efficiency. Figure 11.6a shows an example of how well HICUM can describe the relevant portions of the output characteristics of a power transistor. This along with an accurate description of the charge storage elements (cf. comparison in [27]) results in an accurate prediction of power gain and power added efficiency as shown in Fig. 11.6b.

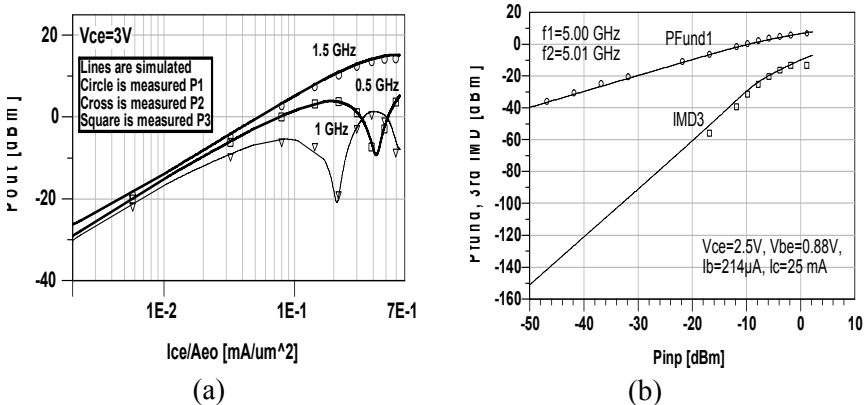


Fig. 11.5: (a) Output power vs. collector current density (at  $V_{CE} = 3\text{V}$ ) for a fundamental frequency of 0.5 GHz and (b) Output power and third-order intermodulation distortion power vs. input power. Comparison between measurement (symbols) and HICUM (lines) [26]. (© 2003 IEEE)

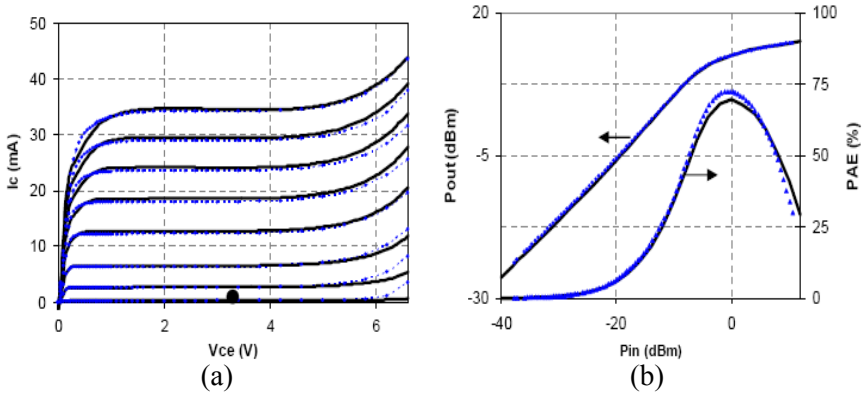


Fig. 11.6: (a) Output characteristics at constant base current and (b) output power and power added efficiency (PAE) vs. input power at 5.75 GHz for a  $3 \times 0.8 \times 20 \mu\text{m}^2$  SiGe HBT. Comparison between measurements (dots) and HICUM (solid lines). The base current values in (a) are (0.01, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5)  $\times 384 \mu\text{A}$ . The impedances were tuned to achieve maximum PAE [27]. (© 2007 IEEE)

An important physical mechanism to be covered for circuit design is noise. Advanced HBTs with a triangular or trapezoidal Ge profile possess a high base drift field, so that the transit time through the BC SCR makes up a significant contribution of the overall  $\tau_f$ . Therefore, the correlation between dynamic base current and transfer current (cf. ch. 7) starts to have a visible impact on the high-frequency noise characteristics. This is shown in Fig. 11.7a for the minimum noise figure  $NF_{min}$ : while at frequencies far below  $f_T$  noise correlation is negligible, it lowers  $NF_{min}$  significantly at frequencies approaching  $f_T^3$  (cf. 26GHz curve) [28]. In contrast, as shown in Fig. 11.7b, for LEC HBTs having an almost box-like Ge profile the correlation is barely visible despite a lower  $f_T$ . HICUM provides an excellent approximation of the measured data not only at low frequencies but also at high frequencies when noise correlation is included.

With increasing transistor speed the breakdown voltage decreases, leading to a higher importance of an accurate description of the collector avalanche process and the associated noise. According to  $NF_{min}$  exhibited in Fig. 11.8a up to high CE voltages, HICUM agrees quite well with experi-

3. Note that not peak  $f_T$  but the bias point value is relevant here.

mental data. For verification purposes also the results of device simulation have been inserted. With the confidence in their accuracy having been established at lower frequencies, HICUM is compared to  $NF_{min}$  from device simulation up to 100 GHz [29], i.e. significantly beyond peak  $f_T$  and the capability of the measurement equipment. This result validates the suitability and accuracy of the model for noise simulations at both high frequencies and high CE voltages.

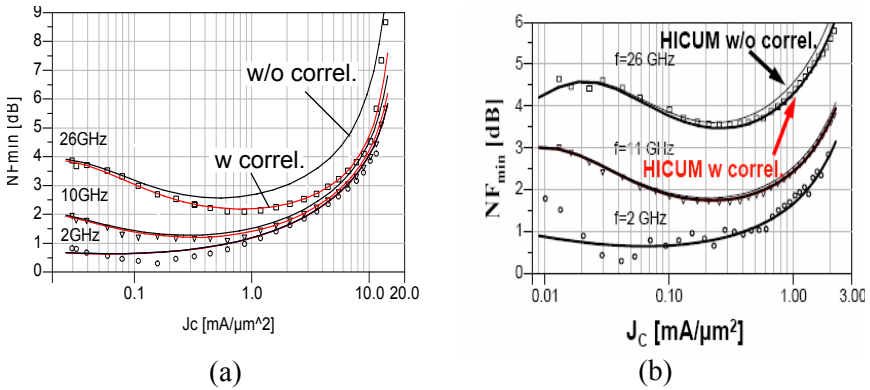


Fig. 11.7: Minimum noise figure vs. collector current density at different frequencies [28]: comparison between measurement (symbols), HICUM with and without correlation. (a) 150 GHz SiGe HBT of JazzSemi and (b) 90 GHz SiGe LEC HBT of Atmel. (© 2006 IEEE)

HICUM has also been successfully applied to III/V HBTs (e.g. [30, 31, 32, 33, 17]). An example was published in [34] for an InGaP/GaAs HBT process. Here, HICUM was specifically selected due to its sophisticated and accurate charge storage description. The only difference to the standard equations of ch. 8 was an extension of  $\tau_{j0}$  by a term  $k_{BE}/I_T^{1/2}$  for the neutral BE charge that follows directly from the ideal result of (3.236) for  $\tau_{BE}$ . Another recent attempt at developing III/V HBT compact models [35, 36] has also directly employed several features of HICUM, especially those related to charge storage.

Temperature dependent characteristics were compared in, e.g., [37, 38]. An extension to describe SOI HBTs is described in [39]. The impact of high current and voltage stress on the reliability behavior of SiGe HBTs was investigated in [40] using HICUM for modeling self-heating and low-

frequency noise and the corresponding interaction. Finally, more examples for comparisons can be found in, e.g., [41, 42, 43, 44, 45, 9, 11, 12]), and a sample parameter list in [46].

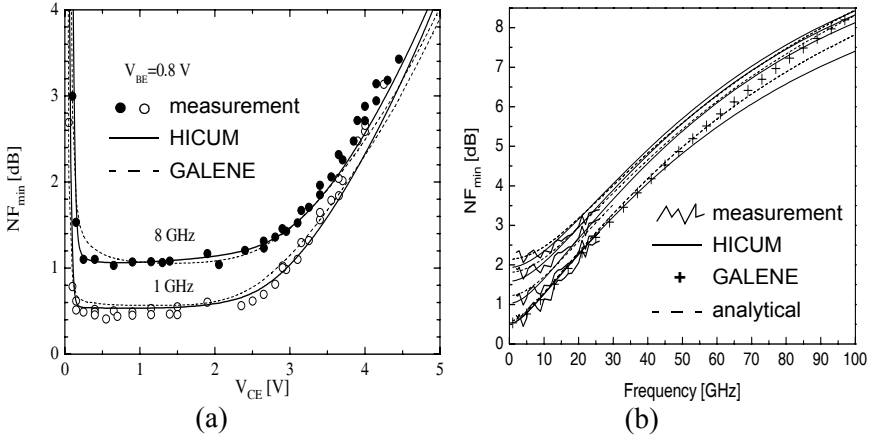


Fig. 11.8: Comparison between measurement, HICUM, and HD device simulation (GALENE) for a power SiGe LEC HBT [29]. (a) Minimum noise figure vs. CE voltage at  $f = (2, 8)$  GHz and (b) minimum noise figure vs. frequency (at  $V_{CE}/V = 2, 3, 3.5, 4$ ). (© 2009 IEEE)

### 11.2.2 HICUM/L0

The first version of this model was successfully applied in [47] to a SiGe HBT in an  $0.18\mu\text{m}$  BiCMOS process. Comparisons of a more mature model version were presented in [48] for four different process technologies fabricated at three different companies. Examples demonstrating the accuracy of the most recent version 1.2 of the model are shown in Fig. 11.9 [49]. The DC current characteristics in (a) show good agreement up to the onset of the high-current region. Since at high injection the model predicts both currents somewhat lower than the measurements (which could be caused, e.g., by too large series resistances) error compensation still leads to an excellent agreement for the current gain in (c) and (d). The transit frequency in (b) is well approximated in comparison to, e.g. the SGPM, and only shows significant deviations at low CE voltages. Deviations in this operating region are expected due to the simplifications made during the derivation of the model equations.

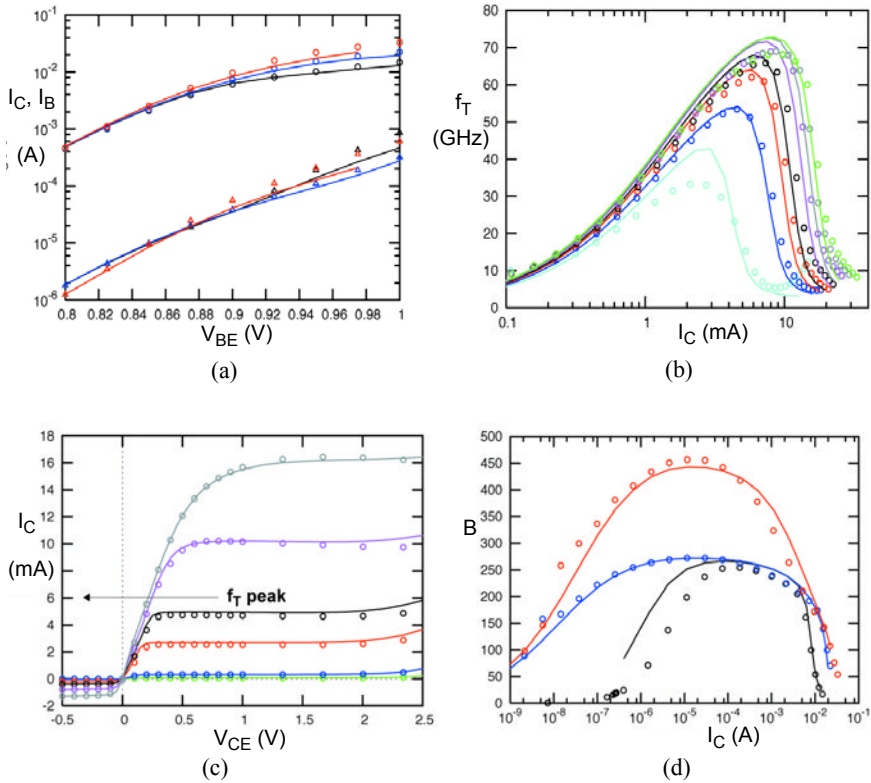


Fig. 11.9: Comparison of HICUM/L0 v1.2 (lines) with experimental results (symbols) of a SiGe HBT fabricated in a  $0.25\mu\text{m}$  BiCMOS process of ST. (a) Gummel characteristics for  $V_{BC}/V = 0.5, 0, 1$ . (b) Transit frequency vs. collector current for  $V_{BC}/V = 0.5, 0.25, 0, -0.25, -0.5, -1, -1.5$ . (c) Output characteristics below and beyond peak  $f_T$ . (d) DC current gain vs. collector current for  $V_{BC}/V = 0.5, 0, -1$ .

### 11.2.3 HICUM/L4

Figure 11.10 exhibits the output characteristics of a 20 GHz SiGe HBT power transistor with the dynamic load circles overlaid. Such load circles are obtained from operating a transistor, biased at a given operating point, with a certain input power at different frequencies. During a single period the collector current as function of the output voltage ( $V_{CE}$ ) starts deviating from the low-frequency load line due to the capacitive and inductive passive elements used power matching and for biasing. With increasing

frequency the resulting load circles shrink, which corresponds to a reduction of output power. This is clearly visible in Fig. 11.10 from the HICUM/L2 results. Since the selected frequencies and its harmonics are fairly large compared to the transit frequency of the investigated transistor, also a distributed electro-thermal HICUM/L4 model was generated directly from the scalable specific parameter set of HICUM/L2 using TRADICA. According to Fig. 11.10 the distributed effects can change the size and form of the load "circle" significantly already at  $f=2.4$  GHz. Further investigations on more examples and regarding the exact cause of such differences are presently being performed.

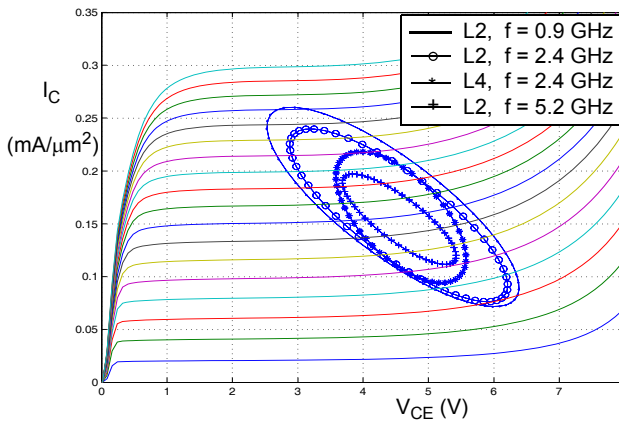


Fig. 11.10: Simulated load circles for  $f/\text{GHz} = 0.9, 2.4, 5.2$ . In addition, a comparison between HICUM/Level2 (o) and HICUM/Level4 (\*) is shown for  $f = 2.4$  GHz.

#### 11.2.4 Statistical and predictive modeling

Today's semiconductor manufacturing business has become very competitive due to the rise of "pure play" foundries. This together with the escalating mask and fabrication cost requires extremely efficient design methods. Among the *necessary* "tools" for achieving this goal are accurate compact models through which the foundry communicates its process capability to the process users, i.e. the circuit design community. The models are also supposed to include information about the unavoidable process variations as to allow centering a design with respect to maximum yield and to trade-off the cost of yield loss versus meeting time-to-market con-

straints. Conventionally, the information on process variations has been provided in form of "worst/best"-case (or generally "corner") parameters. These characterize a single case, in which for instance all parameters are shifted to yield the slowest or fastest possible device performance. Such representation of only two special cases is often not appropriate for the desired application(s), especially for HF analog design. A more adequate approach is to perform random variations of appropriate parameters (based on given distributions) to obtain the statistical distribution of the desired device and circuit related figures of merit (FoMs). For an overview on different methods for statistical modeling and simulation see, e.g., [50] and references therein.

Statistical and yield simulation has been a research focus for quite some time. However, so far commercial design systems offer only very little support for statistical simulation due to several reasons:

- The complexity of the problem requires a significant investment to achieve a generic PDK infrastructure covering all existing process technologies.
- Statistical process data is often very difficult to translate into a description through compact model parameters due to often (w.r.t. modeling) inadequate PCM test structures used by fab engineers, parameter correlations, and non-physical model formulations. Even with suitable statistical process data being available building a reliable statistical representation in the compact model domain usually requires quite some effort and sufficiently physics-based compact models.
- Many foundries consider fab data as proprietary. This causes additional constraints for incorporating such data into PDKs.

While significant and often purely mathematical research work on statistical simulation has focused on MOS technology, only very little effort so far has been dedicated to bipolar technologies and even less to *high-frequency* applications in general.

Typically, statistical simulation offered in circuit simulators is based on random variation of compact model parameters using the Monte Carlo (MC) method. The result of such a simulation, shown in Fig. 11.11a for a 1D SiGe HBT, indicates that the (peak) transit frequency is completely uncorrelated to the low-injection DC collector current. However, performing a random variation of the doping profile as it occurs in a real transistor the corresponding device simulations yield a quite strong correlation as shown

in Fig. 11.11b. This can be easily explained by the transistor theory presented in earlier chapters: the transit time, BE depletion capacitance and transfer current all depend on base doping. Apparently, the standard MC method offered in circuit simulators ignores the correlation between model parameters. The proof is provided in Fig. 11.11c. Here, the dependence of the HICUM model parameters on the underlying common technology parameters, such as base and collector doping or width, have been properly included in the simulation yielding a similar trend as device simulation.

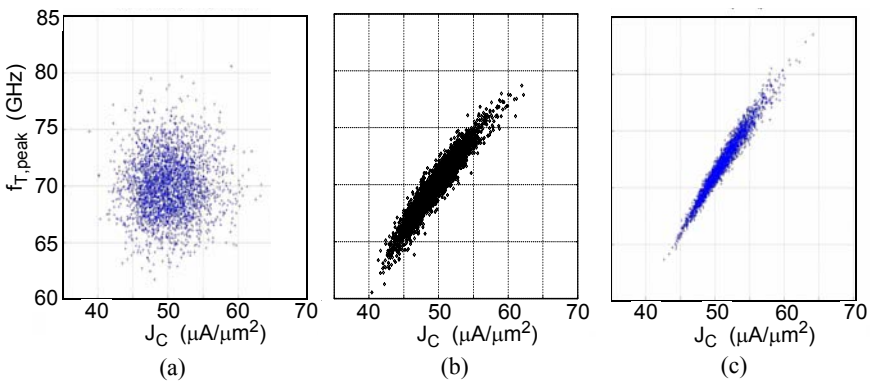


Fig. 11.11: Scatter plots of peak transit frequency vs. low-injection collector current density (taken at *constant*  $V_{BE}$ ) for an intrinsic (i.e. 1D) transistor obtained by (a) MC variation of compact model parameters  $\{c_{I0}, Q_{p0}, C_{JEI0}, C_{JCI0}, \tau_0, r_{CI0}, V_{PT}, \tau_{hcs}\}$  in a commercial circuit simulator, SPECTRE (b) varying the doping profile in 1D device simulation, (c) random variation of technology parameters  $\{N_{Bi}, w_B, N_{Ci}, w_{Ci}\}$  in TRADICA and generation of properly correlated model parameters (see simulation flow in Fig. 11.13).

Obviously, obtaining the correct trend for the statistical variations of FoMs requires the users to include the correlation between model parameters, with their vector being designated below as  $\mathbf{m}$ . For a physics-based model like HICUM many model parameters are already expressed as a function of technology parameters, with their vector being designated as  $\mathbf{t}$ . Assuming  $\mathbf{t}$  as linearly independent fundamental parameters, a given distribution of  $\mathbf{t}$  will then allow the users to obtain the associated set of correlated  $\mathbf{m}$ , from which the transistor and circuit characteristics as well as their distributions can be calculated. The direct application of  $\mathbf{m}(\mathbf{t})$  requires a knowledge of the actual values for  $\mathbf{t}$  or at least their distribution around a



mean value  $t_0$ . This is generally not the case though since it is very difficult and time consuming to measure doping concentrations or junction depths and their process variations. Note that the mean represents the *actual* state of a process at a given time, while the nominal vector represents the *desired target* state.

In practice, *electrical* measurements are employed for obtaining an overview on process stability and variation. This is done by performing process control monitor (PCM) measurements on all wafers in production. In general, a PCM consists of a test structure and its associated measured value. The latter may have to be converted or corrected to obtain the value of a desired variable. The final data vector obtained from a PCM is designated here as  $p$ . Major practical requirements for PCMs are:

- easily measurable in-line;
- strong coupling to important device and model parameters;
- structure size fits into the reserved available space (usually the scribe-channel);
- simple conversion of measured data to meaningful PCM value;
- as weakly correlated as possible among each other for measurement efficiency reasons.

Table 11.1 contains for the internal transistor the list of  $p$  elements,  $p_k$ , and their relation to  $t$  elements,  $t_n$ . Since only deviations from a mean are of interest, the matrix has already been formulated in terms of relative deviations (indicated by a  $\delta$ ) and ratios  $r$  of absolute values. They are defined as

$$\delta v = \frac{\Delta v}{\bar{v}}, \quad r_v = \frac{v^*}{\bar{v}} = \frac{\bar{v} + \Delta v}{\bar{v}} = 1 + \delta v \quad (11.1)$$

with  $\bar{v}$  as reference (mean or nominal) value of a variable ( $p_k$ ,  $t_n$ ),  $v^*$  as varying actual value, and  $\Delta v$  as measured or specified absolute variation. Using ratios is generally preferred, but there are a few exceptions, for which the variations need to be specified as *absolute* values. These include variation of emitter width  $\Delta b_{E0}$  and length  $\Delta l_{E0}$ , of collector thickness  $\Delta w_C$ , and of the bandgap (voltage)  $\Delta V_g$ .

The PCM variables in Table 11.1 correspond to (in the sequence of appearance) the zero-bias internal base sheet resistance, the zero-bias BE depletion capacitance, the zero-bias internal BC depletion capacitance, the low-injection forward collector and base current, and the punch-through

internal BC depletion capacitance. The PCMs  $\delta c_{jEi0}$ ,  $\delta i_{CiI}$ ,  $\delta i_{BiI}$  (i.e. index “i”) represent already specific electrical data that can be obtained directly from large area structures and are preferable from a modeling point of view. However, in advanced SiGe process technologies the profile in structures with sufficiently large emitter area often does not agree anymore with the profile of narrow-emitter structures used for circuit design. Therefore, one has to use the PCM data measured on narrow-emitter structures ( $\delta c_{jE0}$ ,  $\delta i_{CiI}$ ,  $\delta i_{BiI}$ ).

$\delta \mathbf{p} \downarrow \backslash \mathbf{r}_t \rightarrow$	$\mathbf{r}_{NBi}$	$\mathbf{r}_{wB}$	$\mathbf{r}_{NCi}$	$\Delta \mathbf{V}_g$	$\Delta \mathbf{b}_{E0}$	$\mathbf{r}_{JBEiS}$	$\Delta \mathbf{w}_{Ci}$
$\delta \mathbf{r}_{SBi0}$	xxx	xx	(x)	-	(x)	-	-
$\delta c_{jE0}$ $\delta c_{jEi0}$	xxx xxx	—	—	x x	xx -	—	—
$\delta c_{jCi0}$	(x)	-	xxx	x	-	-	-
$\delta i_{CiI}$ $\delta i_{CiI}$	xxx xxx	xx xx	(x) (x)	xxx xxx	xx -	—	—
$\delta i_{BiI}$ $\delta i_{BiI}$	—	—	—	—	xx -	xxx xxx	—
$\delta c_{jCi,PT}$	-	-	-	-	-	-	xxx

Table 11.1: Matrix showing for the internal HBT the fundamental dependence of  $\mathbf{p}$  on  $\mathbf{t}$ , represented here by  $\delta \mathbf{p}$  and  $\mathbf{r}_t$ . The number of crosses indicates the strength of correlation. Parentheses indicate existing but weak dependencies.

The  $\mathbf{t}$  related variables in Table 11.1 have been selected for the following reasons. The internal base region with its average doping concentration  $N_{Bi}$  and width  $w_B$  determines most of the important electrical parameters. Doping concentration  $N_{Ci}$  and width  $w_{Ci}$  of the internal collector region determine the high-current behavior and breakdown characteristics. In HBTs, a change in material composition and, hence, in bandgap (voltage)  $\Delta V_g$  can have a significant impact on the turn-on voltage as well as on the

speed of a transistor. The emitter width (and length) variation  $\Delta b_{E0}$  is usually determined during process development via CD (critical dimension) measurements since no useful direct electrical measurement method exists for a single device. Since various emitter widths are possibly being offered in a process, a normalization for obtaining a relative variation is not suitable here. Finally, the variation of the base current saturation current density  $J_{BEiS}$  allows to model the statistical fluctuation of the (internal) current gain.

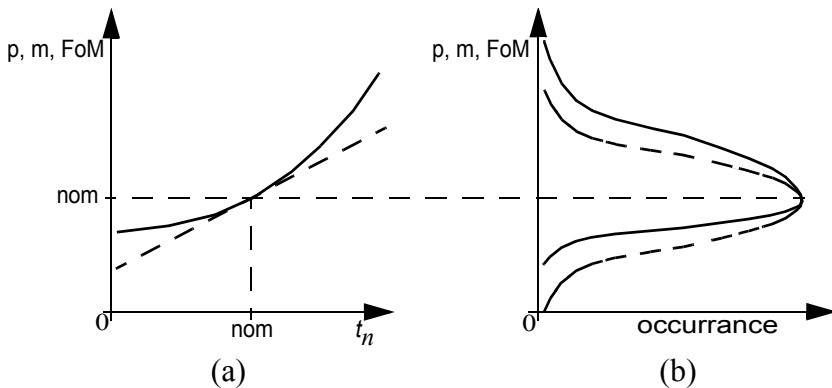


Fig. 11.12: Visualization of the difference in the p.d.f. of a PCM, model parameter, or FoM for a linear (dashed) and nonlinear (solid) relation with  $t_n$ : (a) dependence on  $t_n$ ; (b) resulting p.d.f.

Since most PCM measurements are basically snapshots of selected device parameters or characteristics,  $p$  can be related either directly or through the model parameters to  $t$ . In general, a nonlinear relation exists for both  $m(t)$  and  $p(t)$ , which in turn will cause the probability distribution function (p.d.f.) of a model parameter  $m_j$ , a PCM variable  $p_k$  or an FoM to be "distorted" as visualized in Fig. 11.12. Assuming a normal distribution for  $t_n$  the distribution of  $m_j$  and  $p_k$  will only remain normal for a linear relation (dashed line) in Fig. 11.12a. In contrast, for a nonlinear relation as shown by the solid line, the resulting p.d.f. will be stretched towards larger values and compressed towards smaller values compared to the linear case (cf. Fig. 11.12b). Examples for such a nonlinear dependence on, e.g., doping, are the BE depletion capacitance (of a conventional profile), the saturation currents, and the transit time (i.e. some of its parameters) as well as

the device FoMs (e.g. transit frequency). Therefore, for accurate statistical modeling and design the proper description for  $\mathbf{m}(\mathbf{t})$  and  $\mathbf{p}(\mathbf{t})$  is required.

In many statistical modeling approaches, a *linear* relationship between  $\mathbf{t}$  and  $\mathbf{m}$  or FoMs is assumed a-priori, based on the argument that a statistical analysis reflects "small-signal" changes of a process. This is usually a good approximation for deviations in the order of  $1\sigma$ , but often less appropriate for  $3\sigma$  or  $6\sigma$ , which will require to include the nonlinear relationship.

Based on the general nonlinear analytical relation between  $\mathbf{t}$  and  $\mathbf{p}$  the variation  $\Delta\mathbf{t}$  (or  $\mathbf{r}_t$ ) can be determined from the measured  $\Delta\mathbf{p}$  (or  $\delta\mathbf{p}$ ) as follows. Since the elements of  $\mathbf{t}$ , and thus of  $\mathbf{r}_t$ , are considered linearly independent,  $\delta\mathbf{p}$  can be expressed by a series expansion of  $\mathbf{r}_t$  around its mean

$$\delta\mathbf{p} = \left[ \frac{\partial\delta\mathbf{p}}{\partial\mathbf{r}_t} \right]_{\bar{\mathbf{r}}_t} \mathbf{r}_t + \frac{1}{2} \left[ \frac{\partial^2\delta\mathbf{p}}{\partial\mathbf{r}_t^2} \right]_{\bar{\mathbf{r}}_t} \mathbf{r}_t^2 + \dots \quad (11.2)$$

For sufficiently small variations, i.e. setting  $\mathbf{r}_t$  to its standard deviation  $\sigma_t$ , the series can be truncated after the linear term. Then, the variance  $\sigma_p^2$  of  $\delta\mathbf{p}$  is simply given by (e.g., [51])

$$\sigma_p^2 = \left[ \left( \frac{\partial\delta\mathbf{p}}{\partial\mathbf{r}_t} \right)_{\bar{\mathbf{r}}_t} \right]^2 \sigma_t^2 = [A_{pt}] \sigma_t^2, \quad (11.3)$$

where each element in the matrix  $[A_{pt}]$  corresponds to the squared first derivative in (11.2). Since  $\sigma_p$  is known from PCM measurements, the desired variance  $\sigma_t^2$  of the technology parameters can be calculated by inverting  $[A_{pt}]$ , the rank of which is fairly small. This method of "backwards" calculation of the desired variances was applied first in [52] and is referred to as *backward propagation of variances* (BPV). It is obvious from the dependences indicated in Table 11.1 that the linear equation system (11.3) does not yield a solution for any arbitrary vector  $\sigma_p$ . In other words, for a given set  $\sigma_p$  only a corresponding physics-based set  $\sigma_t$  can be calculated.

Figure 11.13 shows the flow-chart for statistical modeling and simulation. It starts with an extraction of model parameters on devices that behave *typical* w.r.t their electrical characteristics. Usually, these typical model parameters  $\mathbf{m}_T$  are extracted on early wafers, i.e. at a time when statistical information is not available yet. The corresponding PCM data are

given by  $\mathbf{p}_T$ . Once sufficient PCM data are available for statistical evaluation, often already during process qualification, mean values and standard deviations of  $\mathbf{p}$  can be determined. Assuming a sufficiently large sample size, the measured mean values for the  $\mathbf{p}$  elements are a physically meaningful combination that represents the nominal process condition. Therefore, this set can be used to shift the existing typical model parameters to a set of nominal parameters,  $\mathbf{m}_N$ . For this, the difference between the mean of  $\mathbf{p}$  and of  $\mathbf{p}_T$  can be used to calculate a nominal set  $\mathbf{r}_{tN}$  and  $\mathbf{m}_N$  (see discussion on predictive modeling further below).

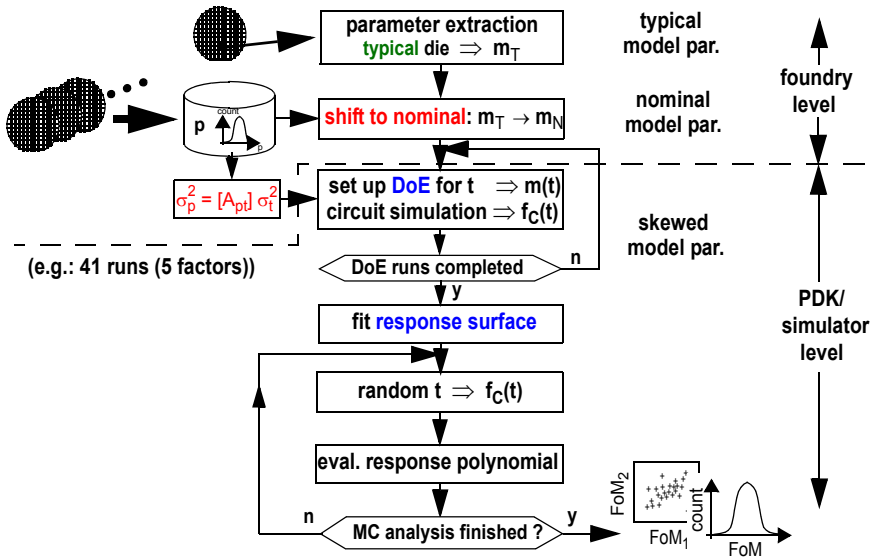


Fig. 11.13: PCM and compact model based statistical modeling and simulation flow (simplified).

The statistical modeling procedure starts with the calculation of the standard deviations  $\sigma_t$  from  $\sigma_p$  according to (11.3). A normal distribution is assumed for the sake of simplicity of statistical modeling and simulation. With  $\sigma_t$  known, vectors of  $\mathbf{r}_t$  can be generated based on the Design of Experiment (DoE) method [53]. For each  $\mathbf{r}_t$  a set of corresponding model parameters  $\mathbf{m}(t) = \mathbf{m}_N + \Delta\mathbf{m}(\mathbf{r}_t)$  is calculated, which are used to obtain the desired FoMs  $\mathbf{f}_C$  from circuit simulation. Note that the length of vector  $\mathbf{f}_C$ , which can include both device and circuit related FoMs, is independent

of that of  $\mathbf{r}_t$ . Once the DoE runs are completed, a response surface is fitted to  $\mathbf{f}_C$  with  $\mathbf{r}_t$  as basis variables. This response surface method allows then to quickly evaluate variations in  $\mathbf{f}_C$  for a large number of random changes in  $\mathbf{r}_t$  so that statistical distributions for  $\mathbf{f}_C$ ,  $\mathbf{p}$ , and  $\mathbf{m}$  can be obtained. It is interesting to note that these data can also be used to find specific combinations of  $\mathbf{r}_t$  that represent certain boundaries for  $\mathbf{f}_C$  such as corner cases.

*Predictive* modeling is just the large-signal version of statistical modeling. While in the latter  $\delta\mathbf{p}(\mathbf{r}_t)$  is linearized, in the former simply the full nonlinear dependence is used. For a given  $\delta\mathbf{p}_T = (\mathbf{p} - \mathbf{p}_T)/\mathbf{p}_T$  a nonlinear equation system results,

$$\mathbf{f}(\mathbf{r}_{t,T}) = \delta\mathbf{p}(\mathbf{r}_{t,T}) - \delta\mathbf{p}_T = 0, \quad (11.4)$$

that is solved iteratively for  $\mathbf{r}_{t,T}$ . Then, the new (shifted) model parameter set can be calculated as  $\mathbf{m}_N = \mathbf{m}_T + \Delta\mathbf{m}(\mathbf{r}_{t,T})$ . Note that the nonlinear equation *system* is fairly small and confined to the variables of the internal transistor. For variables of the external transistor mostly single equations are obtained due to the missing or negligible correlation. Of course, the nonlinear equations can also be evaluated directly for any given vector  $\mathbf{r}_t$  that is obtained from, e.g, process simulation. This way, model parameters and FoMs can be predicted very quickly already early in the process development phase.

Both the predictive and the statistical modeling procedure have been implemented in TRADICA. The set of equations enables not only insight into the effects of process variations on key device parameters but also orders of magnitude faster evaluation of the impact of process variations on electrical device and circuit characteristics than process and device simulation or fitting compact model parameters to measured statistical data. The approach has been applied for production technologies [54]. Implementations of statistical modeling using HICUM have also been reported in [55, 13].

### 11.3 Model deployment for circuit design

Since its selection as one of the standard compact models, HICUM/L2 has become available in the vast majority of circuit simulators such as (in alphabetical order): ADS, AnalogOffice, APLAC, ASX, AVOSpice, EL-

DO, Gemini, GoldenGate, HSPICE, NEXXIM, Ngspice, QUCS, SIMetrix, Smart-SPICE, SPECTRE, SPICE3F, TEKSPICE, UltraSim. This list of simulators and the implemented model versions are always just a snapshot in time, since both model and simulators are constantly being improved. Thus, for regular and latest updates on model availability users are referred to the respective model web-site [56] and simulator vendor contacts.

Since circuit simulators differ in the way they evaluate a model internally and also still do not offer a uniform interface for model implementation, the model code has to be arranged differently in every simulator. Usually, the code is optimized for computational efficiency in the particular simulator. In addition, the derivatives for setting up the Jacobian can become quite complicated in advanced models. This coding effort and the fact that simulator vendors often consider their optimized code as proprietary has made it basically impossible for model developers to directly implement and maintain advanced compact models in more than one circuit simulator. Fortunately, this situation changed a few years ago when the hardware description language Verilog-A [57] became powerful enough to allow advanced transistor models to be implemented. The major advantages of this approach are:

- With a standardized (Verilog-A) syntax the model can be run in all simulators, allowing quick evaluations, tests, and comparisons by users.
- A consequence of the above is that only a single code version needs to be maintained and distributed by model developers, relieving the latter from the heavy burden of coding in unfamiliar simulators.
- Verilog-A does not require the implementation of derivatives, which significantly reduces the coding effort even further and helps speeding up model development.
- The standard syntax also led to the emergence of so-called model compilers [58], that translate the code into a regular programming language (such as C) and a format that fits the actual model integration interface of a particular simulator. This code is then compiled leading to a linked library of models with fast execution time. The continuous improvement of model compilers has also addressed model code optimization for computational efficiency. The whole process of model implementation can now be automated.

- Using model compilers, foundries are able to distribute proprietary compact models as compiled libraries to their customers in order to stay competitive with their PDK offering.

Due to these advantages HICUM/L2 was the first standard transistor model that was distributed in Verilog-A. Although this approach has been a major improvement for model developers there are few issues to be mentioned:

- For most simulators it is still quite difficult to set up the model compiler properly. Therefore, a compiler integrated with the simulator is highly desirable.
- Verilog-A is not as powerful as a regular programming language, making it difficult to implement certain model formulations such as internal iterations, NQS effects, and noise correlation. In most cases adjunct network solutions can be found, which sometimes though appear to be an inferior workaround and make a model look more complicated again. However, since the focus of Verilog-A is not the implementation of compact models, its restrictions may well be justified for other reasons.

Benchmarking of several existing HBT models was performed in [45] regarding accuracy, simulation time and numerical stability. For this investigation, a prescaler (1/8 divider configured by three D flip-flop stages) was used to test free-running transient simulation, a Gilbert-cell downconversion mixer was used for IIP3 simulation through harmonic balance simulation, and a gain stage was used to emulate PA simulation. While the differences between the advanced models were mostly within process and extraction variations, HICUM/L2 appeared to be the best choice from the overall performance point of view, including accuracy and reliable convergence behavior.

Since HICUM/L2 has already been available in circuit simulators for many years, a growing number of semiconductor foundries have been providing model parameters [21]. As a result, the model has been employed in exploratory and production designs fabricated at leading SiGe BiCMOS foundries around the world for already some time. Selected examples are briefly summarized below. In the corresponding figures, “circuit simulation” always refers to simulations “using HICUM”. For process evaluation, often benchmark circuits are designed, which typically include a



current-mode logic (CML) ring oscillator, a low-noise amplifier (LNA), and a power amplifier (PA).

CML and LNA circuits were fabricated in ST's 90-nm 230-GHz self-aligned SiGeC HBT BiCMOS technology [59]. The single-stage SiGeC HBT LNA exhibits an operating frequency range of 15 to 39GHz. As shown in Fig. 11.14 measured results are found to be in excellent agreement with circuit simulation using HICUM.

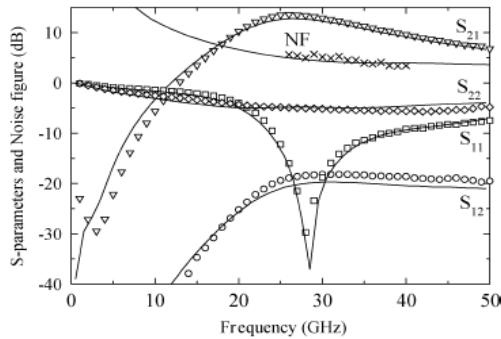


Fig. 11.14: S-parameter results and noise figure for a single-stage LNA [23] fabricated in one of ST's BiCMOS process: comparison between measurement (symbols) and circuit simulation (lines). (© 2005 IEEE)

In ST's 160 GHz 0.13 $\mu$ m SiGe HBT BiCMOS9 production process and subsequent improved variants a number of circuit blocks were successfully designed and fabricated. These include, e.g, an LNA at 24GHz [60]; a +48Gb/s Ethernet retimer [61]; ADCs operating at 40Gb/s and 50Gb/s [62]; an 86Gb/s transmitter including different types of multiplexers [63]; low-power circuits for a 11-to-86 Gb/s serial transmitter [64]; two D-band transceivers featuring an 80-GHz quadrature Colpitts oscillator with differential outputs at 160 GHz, a double-balanced Gilbert-cell mixer, and 170-GHz amplifiers [65]; a low-voltage 77-GHz radar chip set [66], and a single-chip direct-conversion 77–85 GHz transceiver for Doppler radar and millimeter-wave imaging as well as a 86–96 GHz receiver [67], both for automotive applications. An overview on products, mostly for optical communications at 10Gb/s up to 100Gb/s, fabricated with BiCMOS9 and using its model design kit is given in [59].

A set of circuits targeted at wireless communications has also been designed and fabricated in Atmel's 90 GHz SiGe HBT BiCMOS SIGE2 process. These circuit blocks include a high isolation single-pole double-throw differential absorptive switch at 24 GHz for pulsed ultra-wide-band applications [68], a 12-GHz active phase shifter with an integrated low-noise amplifier [69], and a 77 GHz sub-harmonic balanced mixer [70]. The latter circuit is pushing the technology to its limit making accurate modeling crucial for a successful design. In [71], another large set of circuits for a 24 GHz radar system was designed and fabricated in SIGE2. Figure 11.15 exhibits corresponding examples for comparison with simulations.

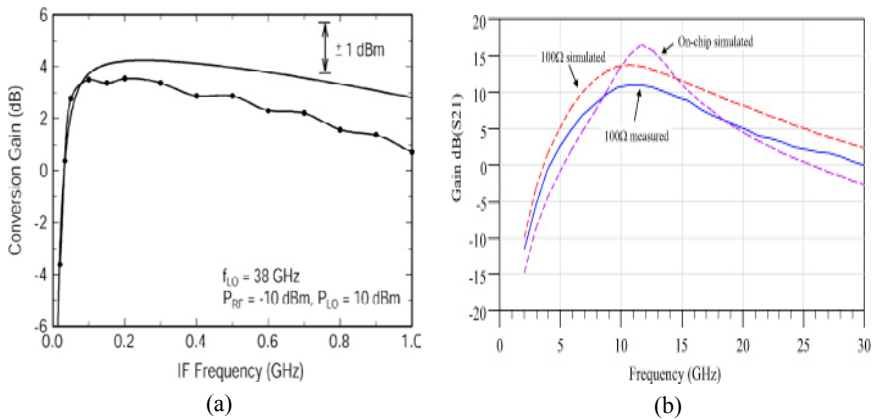


Fig. 11.15: (a) Conversion gain of a 77GHz sub-harmonic mixer [70] (© 2005 IEEE) and (b) high-current LO pump amplifier [71] both fabricated in Atmel's SIGE2 process. Comparison between measurement (symbols) and circuit simulation (lines).

Successful designs have also been fabricated in variants of the 0.18  $\mu\text{m}$  SiGe HBT BiCMOS SBC18 process from Jazz Semiconductor. These include an 8-element phased array receiver for X- and K-band applications [72], a 1:8 differential power divider with a 3-dB bandwidth from 37 to 52 GHz [73], a PHS transceiver [74], a complete DVB-T tuner chip designed at RFMagic [75], a fully differential 40-Gb/s cable driver [76], and a single-chip 65 GHz SiGe radio receiver circuit set (including LNA, transformer balun, downconversion mixer, IF amplifier, and a 65-GHz VCO) [77].

Figure 11.16a exhibits the measured frequency dependence of the power gain for the 8 channels of a 1:8 power divider. The simulation describes the measured data quite well. In Fig. 11.16b measurements of the power gain of a three-stage PA (at 2.4 GHz) are shown for 6 selected dies from wafers that were processed differently. The simulation with nominal parameters follows the average value quite well. This study has been done within the framework of capturing the impact of process tolerances through statistical modeling.

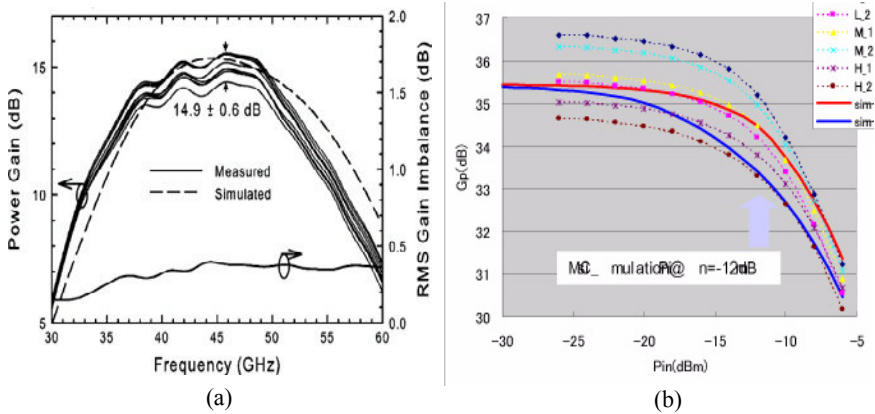


Fig. 11.16: (a) Power gain vs. frequency for a 1:8 power divider [73] fabricated in Jazz's SBC18 process (© 2008 IEEE) and (b) Power gain vs input power for a 3-stage 2.4 GHz PA [55]. Comparison between measurement (symbols) and circuit simulation (lines). The dashed line in (b) corresponds to the results obtained with the SGPM.

Further circuits designed using HICUM include a 60GHz wireless chip set fabricated in IBM's 8HP technology [78], a 80 Gb/s 2:1 multiplexer in a 0.13- $\mu$ m 100 GHz SiGe HBT BiCMOS technology [79], and power amplifiers as well as other circuits for Toshiba's walkman [55]. Finally, the trust in the accuracy of HICUM and its model formulation has also enabled case studies and symbolic modeling based on simulation (e.g [70, 80]).

Above examples represent just a subset of the many circuits that have been designed using the HICUM model and then have been fabricated. The good agreement between experimental and simulation results shown above and in the various publications is achieved, of course, not only by the capabilities of HICUM but also through the thorough parameter extrac-

tion work at the foundries and their partners. Moreover, credit needs to be given also to the circuit designers for performing the sometimes tedious work of the comparisons, which are extremely useful though for both modeling and circuit design. Furthermore, it should be mentioned that the accuracy of the transistor model alone does not guarantee good agreement to circuit measurements but that also an accurate representation of passive devices and parasitic effects is required, especially at very high operating frequencies. Finally, it should be mentioned that the HBT model libraries (with both HICUM and SGPM) in Atmel and certain Jazz process PDKs have been generated using the geometry scaling features of TRADICA.

#### 11.4 References

- [1] H.-M. Rein and M. Möller, "Design considerations for very-high-speed Si-bipolar ICs operating up to 50Gb/s", IEEE J. Solid-State Circuits, Vol. 31, no. 8, pp. 1076-1090, 1996.
- [2] T. Massier, H. Graeb, U. Schlichtmann, "The sizing rules method for CMOS and Bipolar analog integrated circuit synthesis", IEEE Trans. Comp. Aided Des. Integr. Circ. Syst., vol. 27, pp. 2209-2222, 2008.
- [3] Compact Modeling Council (CMC): <http://www.geia.org/CMC>
- [4] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener and A. Koldehoff, "Physics- and process-based bipolar transistor modeling for integrated circuit design", IEEE Journal of Solid-State Circuits, Vol. 34, pp. 1136-1149, 1999.
- [5] R. Ranfft, "Investigation and optimization of subnanosecond bipolar circuits for highly integrated building blocks", Dissertation (in German), Ruhr-University Bochum, 1981.
- [6] M. Schröter, "Transient and small-signal high-frequency simulation of numerical device models embedded in an external circuit", COMPEL, Vol. 10, No. 4, pp. 377-378, 1991.
- [7] C. Jungemann, A.T. Pham, C. Ringhofer, B. Bollhöfer, B. Meinerzhagen, "Stable discretization of the Boltzmann equation based on spherical harmonics, box integration, and a maximum entropy dissipation principle", Journal of Applied Physics, vol. 100, 024502, 2006.
- [8] [Online] <http://www.xmodtech.com/>

- [9] T.-Y. Lee, M. Schröter and M. Racanelli, "A scalable model generation methodology for bipolar transistors for RF IC design", IEEE Bipolar and BiCMOS Circuits and Technology Meeting, Minneapolis, pp. 171-174, 2001.
- [10] M. Schröter, J. Berntgen, W. Kraus, H.-J. Wassener, and H.-J. Strobel, "Scalable HICUM model for SiGe", ECOMPASS Workshop, Bonn (Germany), 17./18. April 2002.
- [11] P. Kempf, "Manufacturing Mixed-Signal SOC", SRC Workshop Orlando, FL, May 17, 2002.
- [12] S. Fregonese, D. Berger, T. Zimmer, C. Maneux, P. Sulima, D. Celi, "Scalable bipolar transistor modeling with HICUM", Proc. Int. Conf. MIXDES, pp. 416-419, 2004.
- [13] R. M. Malladi, K. Newton and M. Schröter, "Development and Design Kit Integration of a Scalable and Statistical High Current Model for Advanced SiGe HBTs", Proc. WCM, International NanoTech Meeting, Boston (MA), pp. 729-734, 2006.
- [14] "Atmel Introduces Innovative HICUM Transistor Model for Its SiGe Technology", Heilbronn, Germany, Bus. Wire Oct. 8, 2001.
- [15] Y. Zimmermann, M. Schröter, P. Zampardi, A. Klimashov, G. Tkachenko, "III-V HBT Modeling, scaling and parameter extraction using TRADICA and HICUM", IEEE Workshop on Power Amplifiers, San Diego, CA, 2003.
- [16] J. Hu, P. J. Zampardi, C. Cismaru, K. Kwok, and Y. Yang, "Physics-Based Scalable Modeling of GaAs HBTs", Proc. BCTM, pp. 176-179, 2007.
- [17] XMOD internal report, 2008.
- [18] B. Ardouin, "Contribution to modeling and characterization of SiGe HBTs at high frequencies", Dissertation (in French), U. Of Bordeaux I, 2001.
- [19] D. Berger, "Study and evaluation of a bipolar transistor model for applications at high frequencies", (in French), Ph.D. thesis, University of Bordeaux, France, 2004.
- [20] F. Blanchet, "Analyse et caractérisation des performances en puissance de transistors bipolaires à hétéro-jonction SiGe:C pour des applications de radiocommunications portables", Dissertation, U. Limoges, France, 2007.

- [21] Production HICUM parameters have been available, as to the best knowledge of the applicant, from following BiCMOS *foundries*: Atmel, IBM, Jazz, ST, Toshiba, TSMC.
- [22] HICUM Workshop, Digest of Presentations, 2001-2009, at [www.tee.tu-dresden.de/tee/eb/](http://www.tee.tu-dresden.de/tee/eb/).
- [23] P. Chevalier, C. Fellous, L. Rubaldo, F. Pourchon, S. Pruvost, R. Beerkens, F. Saguin, N. Zerounian, B. Barbalat, S. Lepilliet, D. Dutartre, D Celi, I Telliez, D Gloria, F Aniel, F Danneville, and A Chantre, "230-GHz Self-Aligned SiGeC HBT for Optical and Millimeter-Wave Applications", IEEE J. Solid State Circuits, vol. 40, pp. 2025-2034, 2005.
- [24] H.-M. Rein and M. Schröter, "A compact physical large-signal model for high-speed bipolar transistors at high current densities - Part II: Two-dimensional model and experimental results", IEEE Trans. Electron Dev., Vol. 34, pp. 1752-1761, 1987.
- [25] M. Schröter, D.R. Pehlke and T.-Y. Lee, "Compact modeling of high-frequency distortion in Si integrated bipolar transistors", IEEE Trans. on Electron Dev., Vol. 47, pp. 1529-1539, 2000.
- [26] P. Sakalas, M. Schröter, W. Kraus, and L. Kornau, "Modeling of SiGe power HBT intermodulation distortion using HICUM", Proc. ESSDERC, Lisboa (Portugal), pp. 311-314, 2003.
- [27] R. Malladi, M. McPartlin, A. Joseph, H. Lafontaine, M. Doherty, "Large Signal Modeling of High Efficiency SiGe HBTs for Power Amplifier Applications", IEEE Proc. Bipolar/BiCMOS Technology Meeting, pp. 58-61, 2007.
- [28] P. Sakalas, A. Chakravorty, J. Herricht, M. Schröter, "Compact Modeling of High Frequency Correlated Noise in HBTs", Proc. Bipolar Circuits and Technology Meeting (BCTM), Maastricht (Belgium), pp. 279-282, 2006.
- [29] P. Sakalas, M. Ramonas, M. Schröter, C. Jungemann, A. Shimukovich and W. Kraus, "Impact Ionization Noise in SiGe HBTs: comparison of device and compact modeling with experimental results", IEEE Trans. Electron Devices, Vol. 56, No. 2, pp. 328-336, 2009.
- [30] P. Sakalas, M. Schröter, P. Zampardi, H. Zirath, and R. Welser, "Microwave Noise Sources in AlGaAs/GaAs HBTs", IEEE International Microwave Symposium, Seattle (WA), pp. 2117-2120, 2002.
- [31] P. Sakalas, M. Schröter, P. Zampardi, H. Zirath, "Shot noise reduction in AlGaAs/GaAs and InP/GaAs based HBTs", Proc. SPIE Fluc-

- tuations and Noise, Santa Fe (NM), eds. J. Deen et al., pp. 387-396, 2003.
- [32] P. Sakalas, M. Schröter, P. Zampardi, M. Racanelli, "Microwave noise in III-V and SiGe-based HBTs: comparison, trends, numbers", (inv. paper), Proc. SPIE 18th Int. Conf. on Noise and Fluctuations, Canary Islands (Spain), pp. 151-163, 2004.
  - [33] J. Herricht, P. Sakalas, M. Schröter, P. Zampardi, Y. Zimmermann, F. Korndorfer, and A. Simukovic, "Verification of  $\pi$ -equivalent circuit based microwave noise model on AIII/BV HBTs with emphasis on HICUM", Proc. International Microwave Symposium, Long Beach, California, pp. 1419-1422, 2005.
  - [34] S.-C. Tseng, C. Meng, W.-Y. Chen, J.-Y. Su, "A modified HICUM model for GaInP/GaAs HBT devices", Microwave and Optical Technology Lett., Vol. 48, No.4, pp. 780-783, 2006.
  - [35] M. Iwamoto, "Linearity characteristics of InGaP/GaAs heterojunction bipolar transistors and power amplifiers", Ph.D thesis, University of California at San Diego, 2003.
  - [36] M. Iwamoto, D. Root, J. Scott, A. Cognata, P. Asbeck, B. Hughes, D. Avanzo, "Large-signal HBT model with improved collector transit time formulation for GaAs and InP technologies", IEEE MTT-Symp. Digest, pp. 635-638, 2003.
  - [37] H. Beckrich, "A contribution to characterizing and modeling of integrated power bipolar transistors in submicrometer BiCMOS technology", (in French), Ph.D. thesis, University of Bordeaux, France, 2006.
  - [38] P. Sakalas, M. Ramonas, M. Schröter, A. Kittlaus, H. Geissler, A. Shimukovitch, "Modeling of SiGe HBT Operation in Extreme Temperature Environment", GHz 2008 Workshop, Chalmers University, Sweden, 1 page, 2008.
  - [39] S. Fregonese, "A compact model for SiGe HBT on thin-film SOI", IEEE Trans. Electron Dev., Vol. 53, pp. 296-303, 2006.
  - [40] M. Diop, M. Marin, N. Revil, G. Ghibaudo, "Reliability Review of Fully Self Aligned 250 GHz Heterojunction Bipolar Transistors for Radio Frequency Applications", IRPS, Paper 2C.3, 2009.
  - [41] U. Basaran and M. Berroth, "Large-signal modeling of SiGe HBTs including a new substrate network extraction method", GaAs Symposium, pp. 437-440, 2003.

- [42] M. Schröter, "High-Frequency Circuit Design Oriented Compact Bipolar Transistor Modeling with HICUM", IEICE Trans. Electron., Vol. E88-C, pp. 1098-1113, 2005.
- [43] M. Schröter, "Advanced compact bipolar transistor models - HICUM", Chapter 8.4 (pp. 807- 823) in "Silicon heterostructure Handbook", ed. by J. Cressler, CRC Press NY, 2005.
- [44] R. Malladi, V. Borich, S. Sweeney, J. Rascoe, K. Newton, S. Venkatadri, J. Yang and S. Chen, "Two-Tone Distortion Modeling for SiGe HBTs Using the High-Current Model", Proc. WCM, International NanoTech Meeting, Boston, pp. 729-734, 2006.
- [45] S. Yoshitomi, "Benchmarking of newly-developed bipolar SPICE models", Proc. Int. Conf. MIXDES, pp. 85-88, 2004.
- [46] R. Malladi, "Sample SiGe HBT compact model parameters", Ch. A4 in "Silicon heterostructure Handbook", ed. by J. Cressler, CRC Press NY, pp. 1201-1210, 2005.
- [47] M. Schröter, S. Lehmann, H. Jiang, and S. Komarow, "HICUM/Level0 – A simplified compact bipolar transistor model," Proc. IEEE BCTM, pp. 112–115, 2002.
- [48] S. Fregonese, S. Lehmann, T. Zimmer, M. Schröter, D. Celi, B. Ardouin, H. regoich, P. Brenner, W. Kraus, "A Computationally Efficient Physics-Based Compact Bipolar Transistor Model for Circuit Design—Part II: Parameter Extraction and Experimental Results", IEEE Trans. Electron Dev., Vol. 53, pp. 287- 295, 2006.
- [49] D. Celi, "HICUM/L0 v1.2 Parameter Extraction and Validation", presentation at the Bipolar Working Group (BAK), Hamburg, 2008.
- [50] M Schröter, H. Wittkopf, W. Kraus, "Statistical modeling of bipolar transistors", Proc. Bipolar Circuits and Technology Meeting (BCTM), Santa Barbara (CA), pp. 54-61, 2005.
- [51] P. Bevington, "Data reduction and error analysis for the physical sciences", McGraw-Hill, NY, 1969.
- [52] C. McAndrew, J. Bates, R. Ida, P. Drennan, "Efficient statistical BJT modeling, why is beta more than  $IC/IB$ ", Proc. BCTM, pp. 28-31, 1997.
- [53] G. Box, "Empirical model building and response surfaces", Wiley&Sons NY, 1987.
- [54] W. Schneider, M. Schröter, W. Kraus, H. Wittkopf, "Statistical simulation of high-frequency bipolar circuits", Proc. DATE07, Nice, pp. 1397-1402, 2007.



- [55] S. Yoshitomi, K. Kawakyu T. Teraguchi, H. Kimijina, K. Yonemura, "Modification of compact bipolar transistor model for DFM Applications", Proc. MIXDES, pp. 125-130, 2006.
- [56] [http://www.iee.et.tu-dresden.de/iee/eb/hic\\_new/hic\\_start.html](http://www.iee.et.tu-dresden.de/iee/eb/hic_new/hic_start.html)
- [57] Verilog-AMS Language Reference Manual, Version 2.2, <http://www.eda.org/verilog-ams>.
- [58] L. Lemaitre, C. McAndrew, S. Hamm, "ADMS - Automatic device model synthesizer", Proc. Custom Integrated Circuits Conf., pp. 27-30, 2002. See also: [www.tiburion-da.com](http://www.tiburion-da.com)
- [59] P. Chevalier, "High-speed SiGe BiCMOS technologies for Ethernet communications from 10Gb/s to 100Gb/s", IEEE BCTM Short Course, Monterey, CA, USA, 2008.
- [60] L. Leyssenne, E. Kerhervé, Y. Deval1, N. Demirel, S. Aloui1, N. Deltimple, D. Belot, H. Ezzedine, "Les Amplificateurs de Puissance Radiofréquences et Micro-ondes : Applications dans le Cadre des Projets Européens MOBILIS et UpperMOS et du Projet RNRT VeLo", presentation at IMS, University of Bordeaux, France, Nov. 2007.
- [61] T.O. Dickson, K. H. K. Yau, T. Chalvatzis, Al. M. Mangan, E. Laskin, R. Beerkens, P. Westergaard, M. Tazlauanu, M.-T. Yang, and S. P. Voinigescu, "The Invariance of Characteristic Current Densities in Nanoscale MOSFETs and Its Impact on Algorithmic Design Methodologies and Design Porting of Si(Ge) (Bi)CMOS High-Speed Building Blocks", IEEE J. of Solid-State Circuits, vol. 41, No. 8, pp. 1830–1845, 2006.
- [62] A. Hart and S.P. Voinigescu, "A 1GHz Bandwidth Low-Pass  $\Sigma\Delta$ -ADC with 20GHz to 50GHz Adjustable Sampling Rate", in RFIC Proceedings, pp. 181–184, 2008.
- [63] S.P. Voinigescu, R. Aroca, T.O. Dickson, S.T. Nicolson, T. Chalvatzis, P. Chevalier, P. Garcia, C. Gamier, B. Sautreuil, "Towards a sub-2.5V, 100-Gb/s Serial Transceiver", in CICC Digest, pp. 471–478, 2007.
- [64] T.O. Dickson and S. P. Voinigescu, "Low-power circuits for a 10.7-to-86 Gb/s serial transmitter in 130-nm SiGe BiCMOS," IEEE Journal of Solid-State Circuits, vol.42, no.10, pp. 2077-2085, 2007.
- [65] E. Laskin, P. Chevalier, A. Chantre, B. Sautreuil and S.P. Voinigescu, "165-GHz Transceiver in SiGe Technology," IEEE Journal of Solid-State Circuits, vol.43, no.5, pp.1087-1100, 2008.

- [66] S.T. Nicolson, K.H.K. Yau, S. Pruvost, V. Danelon, P. Chevalier, P. Garcia, A. Chantre, B. Sautreuil, and S.P. Voinigescu, "A Low-Voltage SiGe BiCMOS 77-GHz Automotive Radar Chipset", *IEEE Trans. MTT*, vol.56, pp. 1092-1104, 2008.
- [67] S. T. Nicolson, P. Chevalier, B. Sautreuil, and S. P. Voinigescu, "Single-Chip W-Band SiGe HBT Transceivers and Receivers for Doppler Radar and Millimeter-Wave Imaging", *IEEE Journal of Solid-State Circuits*, vol. 43, no. 10, pp. 2206-2217, 2008.
- [68] T. M. Hancock and G. M. Rebeiz, "Design and Analysis of a 70-ps SiGe Differential RF Switch", *IEEE Trans. Microwave Theory and Techniques.*, Vol. 53, pp. 2403-2410, 2005.
- [69] T. Hancock, and G. M. Rebeiz, "A 12-GHz SiGe Phase Shifter With Integrated LNA", *IEEE Trans. Microwave Theory and Techniques.*, Vol. 53, pp. 977-983, 2005.
- [70] J.-J. Hung, T. Hancock, and G. M. Rebeiz, "A 77 GHz SiGe Sub-Harmonic Balanced Mixer", *IEEE J. Solid-State Circ.*, vol. 40, pp. 2167-2173, 2005.
- [71] Peter Lindberg, "Radio Frequency Integrated Circuits for 24 GHz Radar Applications", Ph. Dissertation, Uppsala University, Sweden, 2005.
- [72] K.-J. Koh and G. M. Rebeiz, "An X- and Ku-Band 8-Element Phased-Array Receiver in 0.18-um SiGe BiCMOS Technology", *IEEE J. Solid State Circuits*, vol. 43, pp. 1360-1371, 2008.
- [73] J. May, G. Rebeiz, "A 40–50-GHz SiGe 1 : 8 Differential Power Divider Using Shielded Broadside-Coupled Striplines", *IEEE Trans. Microwave Theory and Techniques*, vol. 56, pp. 1575-1581, 2008.
- [74] Microwaves & RF, "PHS transceiver fabricated in Jazz BiCMOS process technology", Feb. 2006.
- [75] P. Mannion, "Startup integrates DVB-T tuner tasks", *EETimes* 2007.
- [76] R. A. Aroca and S. P. Voinigescu, "A Large Swing, 40-Gb/s SiGe BiCMOS Driver with Adjustable Pre-Emphasis for Data Transmission over 75W Coaxial Cable", *IEEE Journal of Solid-State Circuits*, vol.43, no.10, pp. 2177-2186, 2008.
- [77] M. Q. Gordon, T. Yao, and S.P. Voinigescu, "65-GHz Receiver in SiGe BiCMOS Using Monolithic Inductors and Transformers", *Techn. Dig. Si Monolithic Integrated Circuits in RF Systems*, pp. 265-268, 2006.

- [78] U. Pfeiffer, A.-V. Garcia, "Millimeter-Wave design considerations for power amplifiers in a SiGe process technology", *IEEE Trans. Microwave Theory and Techniques*, vol. 54, no. 1, pp. 57-64, 2006.
- [79] Z. Yan, W. Zhigong, and L. Wei, "80 Gb/s 2 : 1 multiplexer in 0.13-um SiGe BiCMOS technology", *J. of Semiconductors*, vol.30, No.2, 4 pages, 2009.
- [80] U. Stuermer, W. John, H. Reichl, "Modellierungskonzept fuer leitungsgebundene Stoerungen in sensorischen Mikrosystemen", *EMV 2004*, pp. 403-410, 2004.

This page is intentionally left blank

## **Chapter 12**

### **Future Trends**

This chapter provides a brief overview on present and future trends for high-speed bipolar transistor technology, including SiGe and III-V HBTs. It is shown that cutting edge bipolar technology has a bright future in emerging millimeter-wave and THz applications and associated markets. This also demonstrates the necessity for a continuous effort in device modeling and compact model development for optimizing circuit and process performance.

## 12.1 Technology challenges

The collapse of the internet bubble and associated decline in fiber-communication related installations as well as the general lack of applications pushing system frequencies as aggressively as Moore's law resulted over the past decade in almost unchanged system operating frequency. During the same period the continuous improvement in HF performance of MOS transistors has led to an increasing displacement of bipolar and III-V technology by RF-CMOS in analog front-end circuitry of *high-volume products*. Although this trend is limited to applications operating in the *low* GHz regime (i.e. up to about 10 GHz), it became a fairly widespread belief that dense digital blocks could be realized together with HF analog front-ends on the *same* chip entirely in *low-cost* (digital) CMOS technology. This perception has a number of flaws that are pointed out next along with some attractive alternative approaches involving the use of bipolar technology.

The design of HF analog front-ends requires special devices such as capacitors, inductors, HF and precision resistors, and pnp bipolar transistors as well as HF suitable metallization and special measures for signal isolation from substrate coupling. In very advanced CMOS also the breakdown voltage of digital transistors is very low so that a special analog MOS transistor structure is typically required. The associated number of masks for these analog HF-specific devices increases the overall mask count and cost significantly beyond that of the digital baseline process. Furthermore, advanced digital CMOS itself is *not* a low-cost technology any more due to the dramatically increasing number of materials and additional masks that are necessary both to eliminate undesired physical effects and to improve the performance. "Low-cost" relates to the minimum achievable cost per transistor or gate fabricated in a given process and exploiting the maximum possible chip size. Hence, advanced digital CMOS is the technology of choice for ultra-high integrated *and* large volume consumer applications. However, this economy of scale fails for products with low to medium *volume* or *integration* resulting in much higher than minimum cost per wafer and die. Also, notice that the perception of CMOS being a low-power technology only applies to logic but *not* to analog HF applications.

According to the ITRS [1] the development of new CMOS process generations faces an increasing number of challenges that are caused by physical, electrical and thermal limitations. Necessary additions such as built-in-test and redundant circuitry for yield enhancements increase complexity and cost but not the overall functionality of a system-on-chip (SoC). Besides rapidly approaching physical and manufacturing limits the fabrication volumes required for the most advanced digital CMOS production technology to achieve a viable return on investment will also approach the maximum available number of customers that can actually afford to buy the corresponding products.

For high-volume products operating in the low-GHz range an RF-SoC in a CMOS process node, which may be 2 generations behind the most advanced node, may often be *economically* more viable and preferable over a possible mixed-technology solution. However, at higher frequencies (e.g. 24, 60, 77 GHz) the use of MOS transistors generally leads to larger footprints and power dissipation compared to bipolar transistors fabricated in the *same* process node. An example is the LNA, in which the high input impedance of the MOS transistor requires a significant increase in device size in order to match the source (e.g. antenna) impedance. This in turn results not only in higher operating current and dissipated power but also in lower speed. Further examples based on practical implementations are given in [2], clearly demonstrating for a 77 GHz application the superiority of a 0.13 $\mu$ m BiCMOS technology (with 230 GHz SiGe HBTs) over a 65nm RF-CMOS process (with 150 GHz MOSFETs).

Figure 12.1a visualizes the available HF performance of subsequent BiCMOS and RF-CMOS process nodes. Obviously, as indicated by the arrows, the *same* HF performance can be achieved with BiCMOS based on an older digital CMOS process that can be *up to 6 generations* behind the leading edge digital CMOS process (as indicated by the arrows). As shown in Fig. 12.1b, while the performance of HF circuit blocks is similar in both processes, the cost for the BiCMOS solution becomes much lower due to the significant depreciation of its digital CMOS baseline technology. In other words, the lower complexity of the older digital process corresponds to a significantly lower initial cost that compensates for the additional mask cost of HBTs and HF passive devices. Hence, as long as the product does not demand an ultra-dense digital CMOS portion the BiCMOS solu-

tion usually turns out to be more cost efficient today and in future generations (see further below).

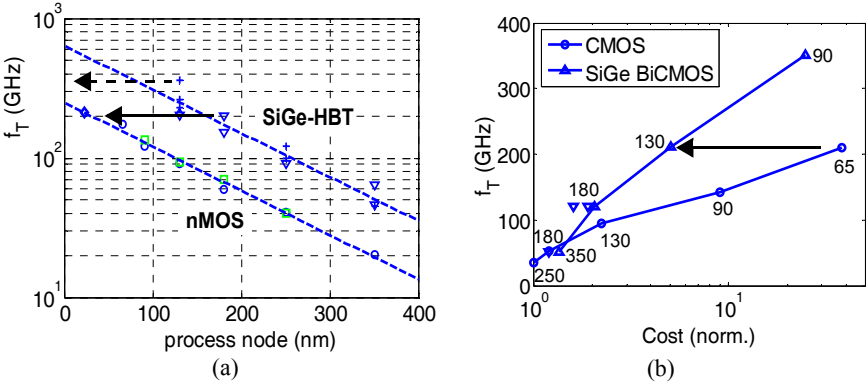


Fig. 12.1: Comparison between CMOS and BiCMOS technology: (a) Trend of SiGe npn HBT and n-MOSFET operating speed (measured by transit frequency  $f_T$ ) over subsequent CMOS process generations indicated by gate length (e.g. [3, 4]). (b) Performance vs. normalized cost for subsequent process generation (indicated in nm by the numbers); points connected with lines correspond to data from IBM [4], separate points to data from other foundries.

From a user and product point of view the functionality of an *RF*-CMOS SoC is still limited and quite far away from the “holy grail” of a complete system on a chip that includes, among others, mechanical sensors, optical interfaces, high-voltage and high-power capability, and very-high-speed data acquisition and processing. Therefore, an alternative approach for implementing such complete systems is to select the process technology that fits best the different tasks of the various portions of the system and then to integrate the different dies into a single package. With the tremendous drop in packaging cost over the past decade, such a mixed-technology system-in-package (SiP) approach allows a much more flexible design and can already today provide a cost-efficient solution, especially at lower volumes. In the long run, the holy grail of a complete SoC could also be a modular fabrication process that allows to flexibly mix different technologies according to product requirements as shown in Fig. 12.2. Obviously, BiCMOS technology serves quite well the demand for more functionality at lower cost *and* low to medium volumes. Due to its inherent advantages for analog HF applications, bipolar and in particular HBT technology is an



excellent candidate for a variety of product areas such as communications, automotive, measurement equipment, entertainment and space applications. Below, a number of examples are presented to illustrate the superiority of BiCMOS over CMOS-only solutions in these areas.

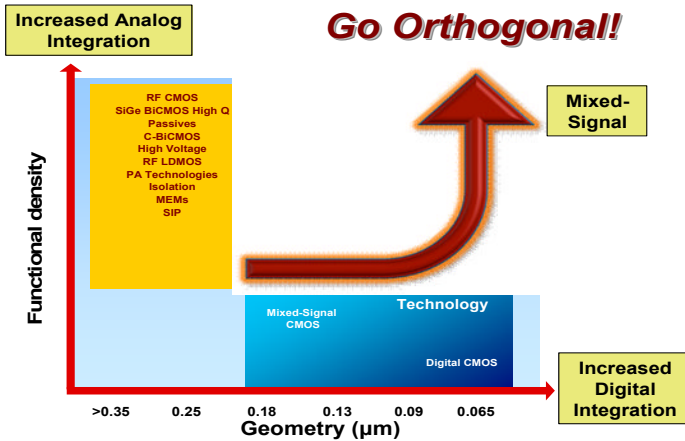


Fig. 12.2: The quest for increased functionality leads to modular integration of diverse device types and associated process technologies [5].

In existing wireless and wireline communications with operating frequencies beyond 10GHz HBT applications comprise front-end circuit blocks. Examples for such blocks include HF low-noise, power and wide-band amplifiers, mixers, laser and interferometer drivers, fast switches, and high-speed/high-bandwidth analog-digital data converters. In hand-sets, GaAs HBTs are dominating the PA market with more than  $10^9$  devices delivered in 2008. SiGe HBTs are attractive for wireless femto-cell base stations, while special high-voltage GaAs HBTs are gaining interest for (short range) pico-cell base stations operating at carrier frequencies around 38 GHz [6]. The ever increasing demand for higher internet access speed and the renewed interest in high-speed residential connections have started to fuel the development of not only backbone fiber-optic circuits operating at data rates beyond 100 Gb/s (e.g. [7, 8]) but also high-speed wireless broadband and point-to-point transceivers operating at 24, 40 and 80 GHz (e.g. [9, 2]) as well as satellite communications operating in the 25 to 40 GHz range [10, 11] and WLANs operating in the 60 GHz band [12]. Fur-

thermore, the increasing demand for high-speed board-to-board data communications between servers and on-chip optical links are other promising applications for high-speed circuits with III-V and SiGe HBT technology.

The most known applications of SiGe HBT technology in automotive applications are radar circuits operating at 24 GHz for side impact and at 77 GHz for front/back collision avoidance (e.g. [13, 14, 15]), which both address the requirements for increased safety. Also, the increasing demand for hybrid and electric cars creates an attractive market for special power semiconductors such as SiC and GaN HBTs (e.g. [16]) that are capable of handling very high power and chip temperature.

The measurement equipment industry has traditionally used HBT and HEMT technology for at least their high-speed front-ends simply because equipment performance needs to be “ahead” of the technology that is to be developed using this equipment. For instance, 50...67 GHz vector network analyzers as well as multi-rate optical/electrical clock recovery modules utilize GaAsSb/InP HBT technology [17, 18]. Furthermore, various types of state-of-the-art equipment, such as scopes, waveform and signal generators, already include SiGe HBT technology [19]. Moreover, SiGe HBTs are excellent candidates for building ultra-high-*bandwidth* ADCs [20]. All applications above have also in common that, at least for the time being, the volume is too low to be attractive for CMOS.

Finally, there is a variety of products requiring linear circuits, power management, high-voltage capability etc.. For instance, BiCMOS technology with “high-voltage” SiGe HBTs dominates the disc drive business.

In all examples mentioned above, the requirement for high frequency performance, high power handling capability, fairly low volume, or the lack of the need for high integration density causes CMOS to be an inferior solution compared to other technologies. Note that the operating frequency numbers given above represent the state of the art in terms of technology at the time this book was written. Naturally these numbers move upward over time, but the reasons for employing different technologies do not change. Regarding the fabrication of analog mixed-signal HF SoCs an interesting and generic observation can be made that is shown in Fig. 12.3. The physical size of the analog portion of an RF-CMOS chip for existing low-GHz applications increases relative to the digital part and can by far dominate the chip size in an advanced, e.g. 65nm, process. This is caused

mostly by (i) the large footprint of passive devices, especially inductors and capacitors, which only shrink with higher frequencies, (ii) the addition of analog circuitry for multi-band solutions, and (iii) of the - with respect to HBTs - large-size HF MOS transistors. In contrast, by using a BiCMOS process a more balanced area ratio is achieved.

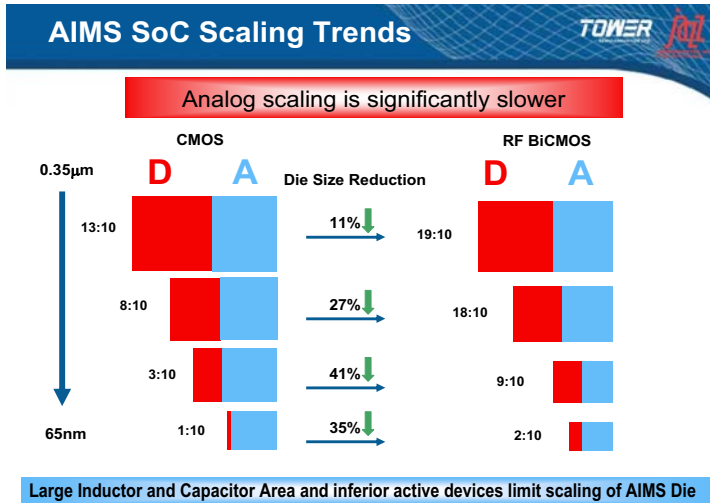


Fig. 12.3: Die size scaling trends for analog mixed-signal HF SoCs [21, 22].

## 12.2 Future applications

Operating frequencies of existing HBTs and HEMTs are already exceeding 300GHz (e.g. [8, 23-28]) and, thus, allow certain HF circuit blocks to operate at 100 GHz and beyond [8, 29, 7, 30, 31]. This is often considered as the lower end of the so-called *THz-gap* the upper limit of which extends to 30 THz. Within the THz-gap a large number of interesting applications are envisioned (e.g. [32, 33, 34]), which are briefly discussed further below. Since this frequency range so far has hardly been tapped into by products and since even the lower frequency range is out of reach for the next several CMOS generations, it is seen as a promising market opportunity for high-performance HBT technologies, including SiGe HBT BiCMOS.

*THz imaging and sensing* is the first category of interest, comprising security, medical, biotechnology, meteorology, material safety, and space applications. Probably the most prominent examples are body scanners for safety control in mass transportation and public buildings. These scanners rely on imaging in the mm-wave range (e.g. 160 GHz band), which allows viewing solid objects through fabric without having a negative impact on the human body as compared to, e.g., x-ray imaging. THz imaging can also be used for (i) chemical detection and analysis of drugs and explosives, (ii) diagnosis of, e.g., DNA, cell and tissue samples in medical and biological sciences, and (iii) reliability testing and diagnosing in material science. Material diagnosis enables, for instance, fault analysis of mechanical devices, of semiconductor VLSI circuits, and of teeth (caries). Earth and space exploration are other areas for THz imaging with applications such as terrain mapping, climate observation, and ozone layer detection. Tera Hertz sensing through radar enables applications such as collision avoidance and intelligent transportation systems, automated landing of airplanes under adverse weather conditions, and planetary entry systems.

A second area for the applications of THz technology are *high-speed/high-bandwidth communications*, ranging from broadband receivers for radio astronomy to data communication front-ends for terrestrial use. Besides fiber-optic circuits operating beyond 120 Gb/s, also high-speed wireless short links are of high interest. The latter enable high-speed downlinks of video data, which is particularly attractive for “in-seat” entertainment in mass transportation systems due to its low weight compared to wireline (fiber) based solutions.

Besides speed, very different requirements have to be met for space applications, in which reliability is key [35]. This also applies to intercontinental communications through deep sea cables, except that the corresponding circuits do not have to be radiation-hard.

### 12.3 Technology trends

Systems with the capabilities listed in the previous section require high-speed/high-frequency components for signal generation and processing. HBT technology in general appears to be a serious contender for building

such systems. Figure 12.4 gives an overview on different HBT technologies with respect to a compromise between speed (measured by  $f_T$ ) and breakdown (measured by open-base breakdown voltage  $BV_{CEO}$ ). At a given speed (i.e.  $f_T$  value) GaN HBTs appear to have the highest breakdown voltage. However, this is an emerging technology for which still many practical obstacles exist. Among the mature technologies, InP HBTs are the most promising candidates, followed by GaAs and SiGe HBTs. In fact InP HBTs have already reached  $f_T$  values beyond 700 GHz (e.g. [23]) but at the expense of relatively poor maximum oscillation frequency  $f_{max}$ . A quite balanced performance was achieved for most advanced InP and SiGe HBT development in [24, 8]. In all cases a trade-off between a circuit relevant base-collector breakdown voltage  $V_{Br,BC}$  and speed ( $f_T, f_{max}$ ) is required that is significantly determined by base and collector doping concentration. The corresponding  $f_T BV_{Br,BC}$  product is roughly constant under the *assumption of drift-diffusion transport* (Johnson limit [36]). However, non-local transport effects can shift this limit to higher values. For instance, once the collector SCR width is too short for carriers to acquire sufficient energy for successful impact ionization (i.e. less than  $\approx 70$  nm in npn SiGe HBTs) the speed may be increased without lowering the BC breakdown voltage. Note though that that  $BV_{CEO}$  does not necessarily limit HF circuit performance [37]. This is also indicated by findings in [38], according to which SiGe HBT circuits yield for mm-wave applications quite similar results as InP HBT circuits.

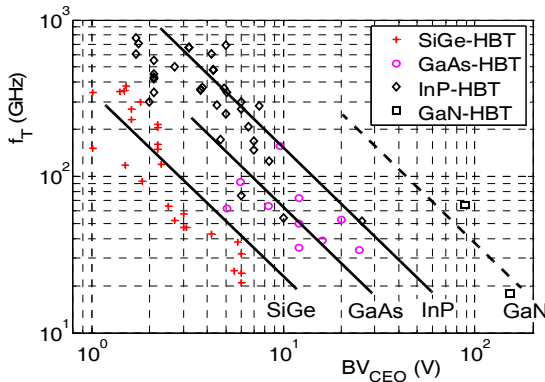


Fig. 12.4: Trade-off between device speed and breakdown voltage for various HBT technologies.

There are presently two larger research projects aimed at aggressively pushing the performance of HBT technology towards mm-wave circuits. A US (DARPA) funded InP process development project presently produces HBTs with  $(f_T, f_{max}) = (560, 560)$  GHz at  $BV_{CEO} \approx 4$  V [8]. The final performance target for the present generation is  $(f_T, f_{max}) = (520, 850)$  GHz, which is expected to yield LNA and PA operating frequencies of up to 430 GHz. The goal is to push InP HBT technology speed by approximately 2013 well into the THz regime with  $(f_T, f_{max}) \approx (1400, 2800)$  GHz [8] and maximum operating frequencies of about 660 GHz for an MS-DFF as well as about 1400 GHz for LNA and PA. The breakdown voltage of such HBTs is not yet known though.

While  $f_T$  and  $BV_{CEO}$  can be measured easily on *devices*, they are not the most representative figures of merit for *circuit* design. Therefore, in the European research project DOTFIVE [39] the CML gate delay  $\tau_{CML}$  (measured with ring-oscillators) and  $f_{max}$  have been selected as device related FoMs for SiGe HBT development. The target values for a 2011 prototype process are  $(f_{max}, \tau_{CML}) \approx (500 \text{ GHz}, 2.5 \text{ ps})$  at  $BV_{CEO} \approx 1.5$  V and moderate  $f_T$  around 400 GHz. The most recent results of  $f_{max} \approx 350$  GHz,  $\tau_{CML} \approx 2.5$  ps, and  $f_T = 300$  GHz at  $BV_{CEO} \approx 1.85$  V [24], all achieved in a  $0.25\mu\text{m}$  lithography, are already very promising. Various options for fabricating ultra-high-speed SiGe HBTs are discussed in [40].

Process development is supported by device simulation. Following the procedure outlined in [41] a doping profile was determined, which yields and partially even exceeds for realistic 3D device structures the target values for  $f_{max}$ ,  $\tau_{CML}$ , and  $BV_{CEO}$  mentioned above. The corresponding profile is shown in Fig. 12.6. The base width is 12.5nm, and the original (epi) collector width is less than 70nm. The lightly doped portion of the emitter prevents excessive tunneling due to the high doping concentration in the base. The Ge profile has a trapezoidal shape with a smaller step at the BE junction to prevent backinjection into the emitter and the associated minority charge storage. The Ge gradient is most important over the distance where the base doping still increases in order to compensate the retarding field. In contrast the large base doping gradient towards the collector already creates a sufficiently large electric field. Note that the overall depth of the active device region does not exceed 100nm so that the drop of the

Ge is not required anymore to be located at the BC junction, but could be moved into the buried layer. However, it is unclear yet whether this would lead to improvements in the electrical behavior.

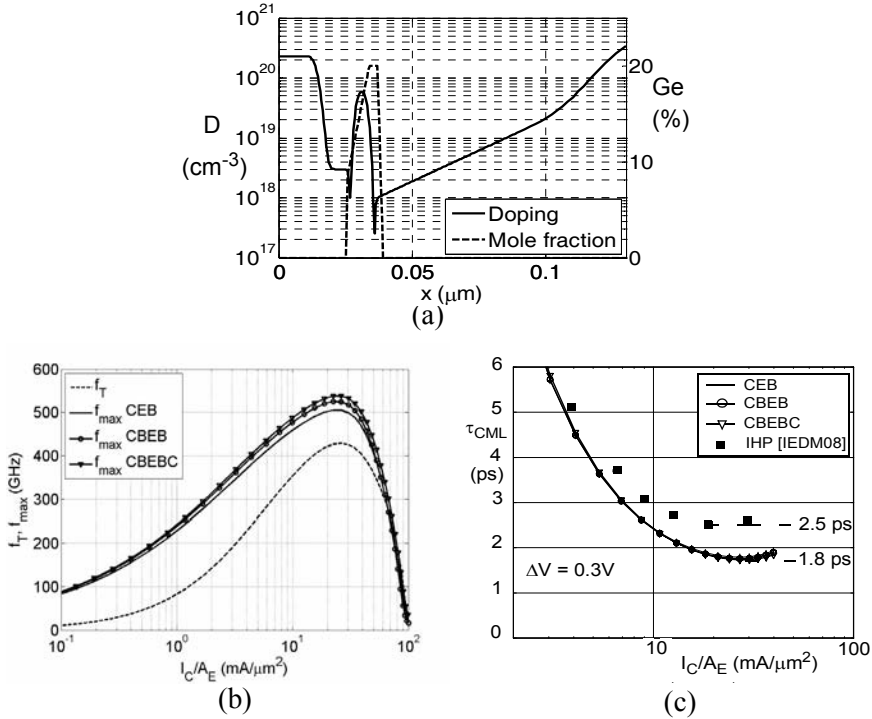


Fig. 12.5: Predicted results for a 500 GHz SiGe HBT. (a) Doping and Ge profile. (b)  $f_T$  and  $f_{max}$  vs. collector current density for different transistor configurations ( $f_{max}$  on-ly) with constant  $A_{E0} = 0.1 \times 0.7 \mu\text{m}^2$ . (c) CML gate delay for different transistor configurations (same  $A_{E0}$ ) with the filled circles indicating measured data from [24].

As Fig. 12.1 demonstrated, even a quite depreciated CMOS process allows extremely fast HBTs, indicating their high performance potential after more aggressive lateral scaling. Figure 12.6 visualizes the expected development of HBT technologies over time. Note that this progress is much more aggressive than the predictions of the ITRS, which appears to be based on applications operating significantly below 100 GHz. With the emerging mm-wave market in sight there is no fundamental reason why

HBT performance can and should not be pushed as aggressively as for CMOS.

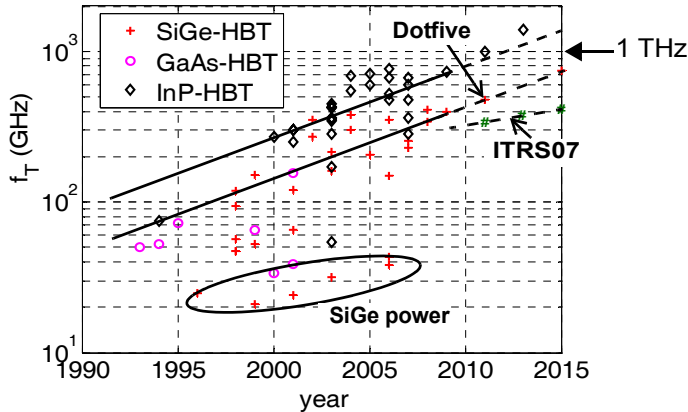


Fig. 12.6: Expected speed of various HBT technologies extrapolated into the near future. For comparison also the ITRS data for SiGe HBTs have been inserted.

Integration of components with multi-100GHz operating frequencies into a larger system is very challenging. If the frequencies at the interfaces between modules can be reduced to a few GHz the SiP approach is feasible. However, this requires a certain integration level for the HF module, which may not be possible with III-V technology. Also, a combination of very-high-speed devices with, e.g., CMOS maybe advantageous for maximizing circuit performance. Therefore, the goal of the DARPA funded COSMOS project [42] is to integrate compound semiconductor devices on finished CMOS wafers at high yields. As proof of concept in the first phase, an InP HBT amplifier has been realized on an existing CMOS wafer with only minimum performance degradation. In later phases the integration of more complex and advanced circuits are planned to follow.

While the above approach enables a much tighter pitch for the interconnect between circuit modules than for a SiP, the ultimate solution would be a modular wafer process architecture that allows to combine a large variety of different device types on the same wafer. This would be a natural extension of already existing analog HF BiCMOS platforms. These “specialty” processes would offer sort of a “plug-and-play” solution for design and system houses, in which only those process modules and associated masks



will be utilized that are required for meeting the electrical specifications and cost of the desired product.

One of these modules may be the underlying digital CMOS process. Generally, as discussed before, this does not have to be fabricated in the most advanced node due to the integration of superior high-speed analog devices. In fact, as Fig. 12.7 shows the mask count (and cost) of RF-CMOS and BiCMOS processes over subsequent generations appears to converge to a level in which the differential between the processes tends to disappear. This makes SiGe BiCMOS even more (cost) competitive besides the fact that SiGe HBTs benefit not only from CMOS lithography but also in a much higher level than MOSFETs from adding Ge. Employing an older CMOS generation, if possible, enables to achieve significantly higher functionality at significantly reduced cost. As a consequence, it is expected that in the future there will be on one hand a number of smaller foundries offering flexible specialty processes as described above and on the other hand a few large foundries and alliances focusing entirely on advanced digital CMOS process development and being capable of shouldering the associated investment. The smaller foundries will be able to cost effectively serve a large variety of products with lower volumes and in application niches.

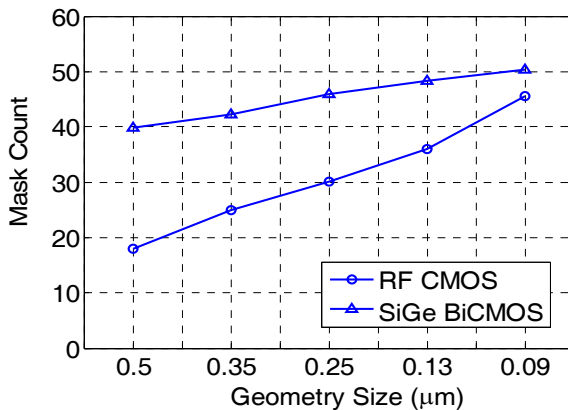


Fig. 12.7: Average mask count of RF-CMOS and BiCMOS over subsequent process generations [43].

## 12.4 Physical limitations and device modeling

In the previous section, circuits operating in the THz-gap have been identified as candidates for driving future development of ultra-high-speed/high-frequency device technologies. Certainly, only the lower end of the THz-gap is expected to be served by electronic devices while the remaining spectrum will require photonic devices. The exact boundary between these will be given by the physical limits of the electronic devices. The most likely factors possibly limiting HBT performance are: low collector breakdown voltage, self-heating, contact resistance, quantum effects, base punch-through, and capacitive parasitics.

With shrinking vertical dimensions the influence of quasi-ballistic transport through the base region (e.g. [44, 46, 47]) and the BC SCR (cf. sec. 4.3.3) increases. These non-local effects need to be taken into account in the model equations for the transfer and avalanche current as well as in the stored minority charge in order to keep the compact model physics-based. On the other hand, shrinking lateral dimensions appear to lead (i) to non-standard scaling with emitter width (e.g. [48], cf. sec. 5.5), (ii) to increased self-heating in deep-trench isolated structures, and (iii) to larger parasitic capacitances due to the close spacing of contacts. Note that in SOI processes, the typical thickness of 0.3...0.4  $\mu\text{m}$  for the buried oxide does not really prevent substrate coupling at very high frequencies but severely increases the device temperature due to self-heating. In extremely scaled devices, tunneling across the BE and BC junction may become significant (e.g. [45]). Depending on how far HBT structures can be scaled while still yielding useful electrical characteristics, carrier confinement due to quantum effects may become relevant in vertical direction (base region) and lateral direction (emitter stack or mesa). The latter “quantum pinch-in” effect reduces the majority carrier concentration towards the lateral surfaces and, hence increases the emitter series resistance. *All these effects need to be taken into account accurately in future compact models in order to avoid limiting both circuit design flexibility and return on investment for process development.*

Compact model development is heavily based on device simulation. While extensive efforts have been spent to tune the physical models of commercial device simulation tools for MOS technology, only little work

has been done, especially for SiGe HBTs. This became obvious during the first phase of the DOTFIVE project [39], which included a work package on simulator calibration. It was clear from the beginning that classical DD based carrier transport is not sufficient anymore for very advanced HBTs. However, the available HD transport models did also yield electrical characteristics that agreed with neither the solution of the BTE, which is considered as reference for the calibration of physical models, nor experimental results. The common observation is that the carrier densities of the BTE solution are better approximated by the DD solution in the base region and by the HD solution in the BC SCR. This results for, e.g., the transit frequency in too low values for DD and too high values for HD. Therefore, the study performed in [49] using default HD transport models yields far too optimistic results. This was confirmed by solving the BTE for the same 1D profile [50, 51]. The corresponding results are shown in Fig. 12.8. Even the calibrated HD simulations in [50, 51] (with slightly different physical models) yield significantly higher  $f_T$  values compared to the BTE solutions. For  $f_T$  predictions, the average of DD and HD results appears to be a good estimate.

In summary, there is still significant work required for developing and providing accurate physical models for HD and (augmented) DD simulation that capture all relevant dependences such as doping, field or carrier temperature, and material composition.

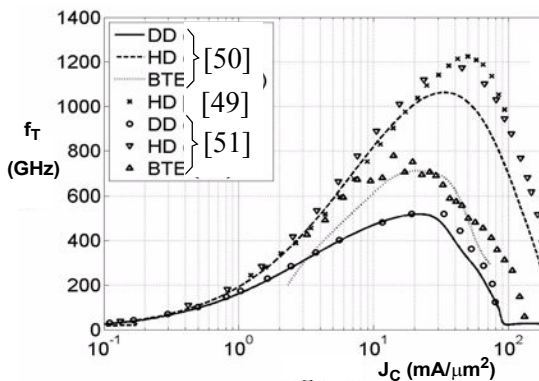


Fig. 12.8: Comparison of 1D DD and HD simulations with results from BTE solutions for the “THz” SiGe HBT profile presented in [49].

## 12.5 References

- [1] International Roadmap for Semiconductors (ITRS), ed. 2007.
- [2] P. Garcia, A. Chantre, S. Pruvost, P. Chevalier, S. Nicolson, D. Roy, S. Voinigescu, C. Garnier, "Will BiCMOS stay competitive for mmW applications?", Proc. IEEE Custom Integrated Circuits Conference, 387-394, 2008.
- [3] Jazz Conference, "Technology update", Newport Beach, CA, Oct. 2008.
- [4] D. Harame, R. Malladi, J. Johnson, J. Pekarik, B. Floyd, J. Dunn, A. Joseph, M. Khater, "Practice and Prospects of Si/SiGe Technology, Applications and Modeling", *Workshop on Compact Modeling for RF/Microwave Applications*, Boston (MA), USA, October 2007.
- [5] G. Amelio, "Technology convergence creating new opportunities for innovation", Plenary talk at IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM), 2008.
- [6] D. Kimball, M. Kwak, P. Draxler, J. Jeong, C. Hsia, C. Steinbeiser, T. Landon, O. Krutko, L. Larson, P. Asbeck, "High efficiency WCDMA envelope tracking base-station amplifier implemented with GaAs HV-HBTs", IEEE Comp. Semicond. Integr. Circ. Symp. (CSICS), pp. 61-64, 2008.
- [7] M. Möller, "Challenges in cell-based design of very-high-speed Si-bipolar ICs at 100Gb/s", Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM), pp. 106-113, 2007.
- [8] M. Rodwell, E. Lobisser, M. Wistey, V. Jain, A. Baraskar, E. Lind, J. Koo, Z. Griffith, J. Hacker, M. Urtega, D. Mensa, R. Pierson, B. Brar, "THz bipolar transistor circuits: Technical feasibility, technology development, integrated circuit results", Proc. IEEE CSICS, pp. 3-5, 2008.  
See also: Technology for Frequency Agile Digitally Synthesized Transmitters (TFAST), US DARPA Project, <http://www.darpa.mil/mto/programs/tfast/index.html>
- [9] H. Izadpanah, "A millimeter-wave broadband wireless access technology demonstrator for the next generation internet network reach extension", IEEE Commun. Mag., pp. 140-145, Sept. 2001.
- [10] C. Vaucher, O. Apeldoorn, M. Apostolidou, J. Dekkers, A. Farrugia, H. Gul, N. Kramer, L. Praamsma, "Silicon-Germanium ICs for satellite microwave front-ends", Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM), pp. 196-203, 2005.

- [11] K.-J. Koh, G. Rebeiz, "A Q-band phased-array front-end with integrated wilkinson couplers for linear power combining in SiGe BiCMOS", *Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM)*, pp. 190-193, 2008.
- [12] S. Reynolds, B. Floyd, U. Pfeiffer, T. Beukema, J. Grzyb, C. Haymes, B. Gaucher, and M. Soyuer, "Silicon 60-GHz receiver and transmitter chipset for broadband communications", *IEEE J. Solid-State Circuits*, Vol. 41, pp. 2820–2830, 2006.
- [13] S. Pruvost et al, "Low Noise Low Cost Solutions for Pulsed Automotive Radar sensors", *Proc. IEEE RFIC*, Session RTUI1B-3, pp. 387-390, 2007.
- [14] S. Nicolson et al, "A Low-Voltage SiGe BiCMOS 77-GHz Automotive Radar Chipset", *IEEE Trans. Microwave Techniques and Theorie*, in press, 2008.
- [15] H. P. Forstner, H. D. Wohlmuth, H. Knapp, C. Gamsjäger, J. Böck, T. Meister, K. Aufinger, "A 19GHz DRO Downconverter MMIC for 77GHz Automotive Radar Frontends in a SiGe Bipolar Production Technology", *Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM)*, pp. 117-120, 2008.
- [16] M. Domeij, M. Nawaz, "SiC BJTs promise to cut the cost of greener driving", *Compound Semiconductors*, pp. 14-16, July 2008.
- [17] Agilent Measurement J., pp. 74-77, Q1 2007.
- [18] H. Liu, O. Ostinelli, Y. Zheng, C. Bolognesi, "600 GHz InP/GaAsSb/InP DHBTS grown by MOCVD with a Ga(As,Sb) graded base and  $f_{TxBV_{CEO}} > 2.5$  THz-V at room temperature", *IEEE IEDM Tech. Digest*, pp. 667-670, 2007.
- [19] see Tektronix website, e.g, [http://www.eetasia.com/ART\\_8800354408.HTM](http://www.eetasia.com/ART_8800354408.HTM); <http://www2.tek.com/cmswpt/prdetails.lottr?ct=PR&cs=nwr&ci=5583&lc=EN>
- [20] Muellrich et al., "DA and AD Converters for 25 GS/s and Above", *IEEE LEOS Summer Topicals*, July 21-23, 2008.
- [21] S. Chaudhury, Jazz Semiconductor, private communication, 2008.
- [22] A. Pawlikiewicz, D. Hess, "Choosing RF CMOS or SiGe BiCMOS in mixed-signal design", *www.rfdesign.com*, pp. 36-44, March 2006.
- [23] W Hafez, W. Snodgrass, M. Feng, "12.5 nm base pseudomorphic heterojunction bipolar transistors achieving  $f_T=710\text{GHz}$  and  $f_{\text{max}}=340\text{GHz}$ ", *Appl. Phys. Lett.*, Vol. 87, 252109, 2005.

- [24] A. Fox, B. Heinemann, R. Barth, B. Bolze, J. Frews, U. Haak, D. Knoll, B. Kuck, R. Kurps, A. Marcschmeyer, H. Richter, H. Rucker, P. Schley, D. Schmidt, B. Tillack, G. Weidner, C. Wipf, D. Wolansky, Y. Yamamoto, "Sige HBT module with 2.5ps gate delay", IEDM Digest, pp. 731-734, 2008.
- [25] M. Khater, J. -S. Rieh, T. Adam, A. Chinthakindi, J. Johnson, R. Krishnasamy, M. Meghelli, F. Pagette, D. Sanderson, C. Schnabel, K. T. Schonenberg, P. Smith, K. Stein, A. Stricker, S. -J. Jeng, D. Ahlgren, and G. Freeman, "SiGe HBT Technology with  $f_{max}/fT = 350/300$  GHz and Gate Delay Below 3.3 ps", IEDM 2004.
- [26] E. Preisler, L. Lanzarotti, P. Hurwitz, M. Racanelli, "Demonstration of a 270 GHz  $fT$  SiGe-C HBTs within manufacturaing-proven  $0.18\mu\text{m}$  BiCMOS process without the use of a raised extrinsic base", Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM), pp. 125-128, 2008.
- [27] P. Geynet, P. Chevalier, B. Vandelle, F. Bossard, N. Zerounian, M. Buczko, D. Gloria, F. Aniel, G. Dambrine, F. Danneville, D. Dutartre, A. Chantre, "SiGe HBTs featuring  $fT > 400$  GHz at room temperature", Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM), pp. 121-124, 2008.
- [28] D.-H. Kim, J. del Alamo, "30nm E-mode InAs PHEMTs for THz and future logic applications", IEDM Digest, pp. 719-722, 2008.
- [29] U. Pfeiffer, E. Ojefors, A. Lissauskas, H. Roskos, "Opportunities for silicon at mmWave and Terahertz frequencies", Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM), pp. 149-156, 2008.
- [30] E. Laskin, P. Chevalier, A. Chantre, B. Sautreuil, S. Voinigescu, "80/160-GHz transceiver and 140-GHz amplifier in SiGe technology", IEEE Radio Frequency Integrated Circuits (RFIC) Symp., pp. 153-156, 2007.
- [31] A. Müller, M. Thiel, H. Irion, H.-O. Ruoss, "A 122 GHz SiGe active subharmonic mixer," in *13th GAAS Symp*, Paris, France, pp. 57-60, 2005.
- [32] M. Tonouchi, "Cutting-edge terahertz technology", Nature Photonics, Vol. 1, pp. 97-105, 2007.
- [33] P. de Maagt, P. H. Bolivar, C. Mann, "Terahertz science, engineering and systems – from space to earth applications," Encyclopedia of RF and Microwave Engineering, Ed. by K. Chang, John Wiley & Sons, Inc., ISBN 0-471-27053-9, pp. 5175-5194, 2005.

- [34] D. L. Woolard, J. O. Jensen, R. J. Hwu, M. S. Shur, "Terahertz Science and Technology for Military and Security Applications", World Scientific, 2007.
- [35] D. Sunderland, J. Lerma, A. McKay, "Designing electronic systems for space", Proc. IEEE BCTM, pp. 171-177, 2008.
- [36] E.O. Johnson, "Physical limitations on frequency and power parameters of transistors", RCA Review, pp. 163-177, 1965.
- [37] M. Rickelt, H.-M. Rein, and E. Rose, "Influence of impact-ionization-induced instabilities on maximum usable output voltage of silicon bipolar transistors", IEEE Trans. Electron Dev., Vol. 48, pp. 774-793, 2001.
- [38] S. P. Voinigescu, T. O. Dickson, R. Beerkens, I. Khalid, P. Westergaard, "A comparison of Si CMOS, SiGe BiCMOS, and InP HBT technologies for high-speed and millimeter-wave ICs", Proc. Top. Meet. on Silicon Monolithic Integr. Circ. in RF Syst., pp. 111-114, 2004.
- [39] "EU project targets 0.5-THz SiGe bipolar transistor", *EE Times Europe* print edition covering March 17 – April 6, 2008. See also DOT-FIVE website <http://www.dotfive.eu/>
- [40] S. Decoutere, S. Van Huynenbroeck, B. Heinemann, A. Fox, P. Chevalier, A. Chantre, T. Meister, K. Aufinger, M. Schröter, "Advanced Process Modules and Architectures for Half-Terahertz SiGe:C HBTs" (inv. paper), IEEE BCTM, pp. 9-16, 2009.
- [41] A. Pawlak, M. Schröter, J. Krause, G. Wedel, C. Jungemann, "On the feasibility of 500 GHz Silicon-Germanium HBTs", Proc. SIS-PAD, San Diego, pp. 27-30, 2009.
- [42] M. Rosker, V. Greanya, T-H. Chang, "The DARPA compound semiconductor materials on silicon (COSMOS) program", Proc. IEEE Comp. Semicond. Integr. Circ. Symp. (CSICS), pp. 6-9, 2008.
- [43] D. Hess, "SiGe BiCMOS or RF CMOS for your next wireless application?", Atmel Applic. J., pp. 27-28, Winter 2006.
- [44] A. Grinberg, S. Luryi, "Diffusion in a short base", Solid-State Electronics, Vol. 35, pp. 1299-1309, 1992.
- [45] J. del Alamo and R. Swanson, "Forward-bias tunneling: a limitation to bipolar device scaling", IEEE Electron Dev. Lett., Vol. 7, No. 11, pp. 629-631, 1986.
- [46] M.A. Alam, M. Schröter, M.S. Lundstrom, "Influence of quasi-ballistic base transport on the small-signal y-parameters of Si bipolar

- transistors”, IEEE Electron Dev. Lett., Vol. 17, Nr. 4, pp. 184-186, 1996.
- [47] M. Vaidynathan, D. Pulfrey, “Effects of quasi-ballistic base transport on the high-frequency characteristics of bipolar transistors”, IEEE Trans. Electron Dev., Vol. 44, pp. 618-626, 1997.
- [48] M. Schröter, S. Lehmann, D. Celi, “Non-standard geometry scaling effects in high-frequency SiGe bipolar transistors”, Proc. WCM, International NanoTech Meeting, Santa Clara, pp. 603-608, 2007.
- [49] Y. Shi, G. Niu, “Vertical profile design and transit time analysis of nano-scale SiGe HBTs for Terahertz  $f_T$ ”, Proc. IEEE Bipolar/BiCMOS Circ. Technol. Meet. (BCTM), 213-216, 2004.
- [50] A. Pawlak, “A HICUM Model for 500 GHz SiGe Heterojunction Bipolar Transistors” (in German), MSEE (Dipl-Ing.) thesis, Chair for Electron Devices and Integrated Circuits, TU Dresden, 2008.
- [51] C. Jungemann, “Physics-based device simulation”, DOTFIVE Report 2008. See also: MONJU User’s Guide V. 1.6.5, ITEM, TU Braunschweig, 2006.



# Index

## A

- analog 22, 723
- application 681, 723
- autocorrelation 476
- avalanche 453, 529,
  - current 115, 655, 118
  - effect 118
  - pinch-in effect 400

## B

- back injection 106, 451, 110
- band diagram 180
- bandgap 622, 22, 187, 434, 449
  - reduction 184
  - variations 258
  - voltage 547
- band-to-band 245
- bandwidth 23
- base contacts 396
- base current 235, 356, 654
- base resistance 363, 374, 384
- base-collector barrier 193
- base-collector breakdown 115
- base-collector depletion charge 251
- base-collector depletion region 264
- base-collector junction 640

- base-collector space-charge region 495
- base-emitter depletion region 267
- base-emitter isolation capacitance 406
- base-emitter junction 639
- base-emitter space charge region 146
- BC capacitance 617
- BC depletion 598
- BC depletion capacitance 119
- BC heterojunction 197
- BC junction 609
- BC space-charge region 155, 169
- BE depletion 600
- BE junction 608
- BE SCR 62
- BE space-charge region 159, 171
- BE spacer 363
- BEBC 683
- bias point 55
- BiCMOS 22
  - process 686
- BJTs 21

- Boltzmann transport equation 35
- boundaries 60
- breakdown 615, 529, 624
- BTB tunnelling 245
- BTE 35
- built-in voltage 442
- butted-collector layout 683

## C

- C code 25
- capacitances 508, 547
- Cardani 650
- CED HBT 186, 196, 213
- characteristics 93
- charge storage 57, 69, 247, 359
- circuit design 36, 702
- circuit optimization 36
- circuit simulators 24
- classical approach 121
- CMC 25, 682
- CML 23
- CMOS 21
  - process 729
- collector current 606, 613
- collector current spreading 536
- collector region 152, 160
- collector-substrate capacitance 413
- collector-substrate junction 544, 642
- commercial simulators 25
- compact formulation 232
- compact model 23, 53
- compact modeling 288
- compilers 25
- complete transistor 535

- complexity 35, 54
- conductivity modulation 274, 314, 531
- contact 434
- correlation 698
- critical current 76, 154
- CS depletion capacitance 601
- current crowding 308, 531
- current dependence 251
- current gain 453
- current spreading 372
- CV measurements 587
- CV meters 587

## D

- DC characteristics 596
- DD 35
- de-embedded S-parameter 597
- deep trench isolation 461
- depletion capacitance 596
- depletion charge 226, 247, 360, 440, 508, 547, 639
- derivatives 703
- design kits 683
- DEVICE 584
- device configuration 56
- device modeling 35
- device sizing 23, 36
- diffusion length 107
- digital 23
- dimensions 56
- distributed effects 635, 694
- distributed model 635, 673
- doping concentrations 56
- doping profiles 55
- double base contact 385
- double-collector layout 683
- drift-diffusion 35

- DTI 461
- dynamic 57
- dynamic emitter current crowding 533
- E
  - EC 55, 304, 505
  - EDA 25
  - EDA tools 686
  - EDA vendors 686
  - EKV3 686
  - ELDO 25
  - electric field 117, 201
  - electrical measurements 55
  - electrical parameters 56
  - electro-thermal coupling 458, 463
  - electrothermal modeling 25
  - emitter current crowding 619
  - emitter length 582
  - emitter perimeter region 540
  - emitter resistance 411, 604
  - emitter window width 582
  - energy gap 22, 547, 622
  - equilibrium 180, 188
  - equivalent circuit 54, 304, 638
  - experimental data 672
  - experimental verifications 25
  - external base resistance 603
  - external base-collector region 542
  - external collector resistance 407
  - external collector resistance 605
  - external series resistances 543
  - extraction flow 592
  - extraction procedure 594
- F
  - Fermi level 179
  - Fermi-potential 188
  - fiber-optic 25
  - field calculations 85
  - flicker 485, 498
  - fluctuating variable 475
  - FoMs 726
  - forward bias 125
  - Fourier transform 476
  - frequency domain solution 281
  - $f_T$  70
  - fully depleted collector 86
- G
  - GaAs 21
  - GaN 21, 725
  - gate delay 23
  - Ge profile 186
  - generalized integral charge-control relation 219
  - geometry 303
    - scaling 208, 682
  - GICCR 208, 275, 449, 507
  - GICCR "master" equation 209
  - Green's function 489
  - g-r-t 483
- H
  - HBT 21, 178
  - HBT circuits 725
  - HD 35
  - heterojunction diode 182
  - heterojunctions 177
  - HF BiCMOS platforms 728
  - HF noise 535
  - HF resistors 686
  - HICUM 24, 681
  - HICUM/L0 628, 634, 692

- HICUM/L2 26, 507, 535, 581, 628, 634, 671, 687
  - HICUM/L4 627, 635, 673, 693
  - HICUM/Level0 636
  - HICUM/Level2 26, 505
  - hierarchy 26, 633, 636
  - high collector voltage 154, 161
  - high current densities 76, 159
  - high injection 101
  - High-CURrent Model 24
  - high-frequency 22
    - effects 616
  - high-injection solution 274
  - high-performance 584
  - high-speed applications 57
  - high-speed circuits 24
  - high-voltage 189, 584
  - hole charge partitioning 225
  - homojunction 75
  - HP 584
  - HV 584
  - hydro-dynamic 35
- I**
- $I_{CK}$  76
  - IC 22
  - III/V HBT 21, 25, 685
  - IMD3 689
  - impact ionization 115
    - noise 496
  - impedance matching 582
  - implementation 24
  - injection width 261, 372
  - InP 21, 725
  - integral charge-control relation 208, 213
  - integrated circuits 22
  - interface 106, 434
  - internal base 602
    - resistance 304, 307, 379, 531
    - sheet resistance 584
  - collector junction 510
  - emitter junction 508
  - internal transistor model 506
  - intra-device substrate coupling 418, 618
  - ionization integral 116
  - ionization rate 117
  - isolation capacitance 598, 600
  - ITRS 727
  - IV measurements 587
- J**
- Johnson limit 725
  - junction current 552
  - junction depth 62
  - junction temperature 55
- K**
- Kelvin structures 587
  - Kirk-effect 85
- L**
- L2 26
  - Laplace transformation 289
  - large-signal model 637
  - large-signal transient 326, 384, 534
  - lateral non-quasi-static effect 533
  - layout 303
  - LCR 587
  - LEC 186
  - Level0 636
  - linearization 281
  - LNA 23
  - low collector voltage 153, 160

- low current densities 150
- low injection 95, 130
- low-voltage 191
- lumped 304
- M
  - matching analysis 36
  - material composition 247
  - maximum oscillation frequency 23
  - MC 696
  - measurement 69, 587, 592
    - conditions 589
  - medium current densities 150
  - medium injection 104
  - metallurgical junctions 63
  - Miller-effect 121
  - millimeter-wave 717
  - MIM 686
  - minimum noise figure 690
  - minority charge 57, 59, 230, 361, 367, 443, 513, 550, 642
  - minority transit time 64
  - mixed-signal 723
  - mobile charge 129, 256
  - mobility 434
  - model comparisons 687
  - model deployment 702
  - model libraries 684
  - model parameter 566, 696
  - Moll-Ross 258
    - relation 212
  - mono-silicon 106
  - MOS 27
  - MOSFETs 22
  - multi-dimensional 208
    - effects 497
  - multiplication factor 115, 118
  - multi-stripe structure 683
- N
  - network analyzer 270
  - neutral base region 131, 151, 167, 257, 487
  - neutral collector 260
  - neutral emitter 142, 264, 495, 156, 171
  - neutral region 59, 178
  - noise 23, 456, 473
    - figure 23
    - mechanisms 474
    - model 561, 664
      - power spectral density 478
  - non-local base-collector avalanche 238
  - non-quasi-static effects 268, 525
  - non-standard-scaling 421
  - NQS effect 274, 535
  - number of nodes 54
  - Nyquist 479
- O
  - optimization task 23
- P
  - parameter 581, 665
    - analyzer 587
    - determination 581
    - extraction 25, 54, 593, 671, 672
      - extraction effort 634
  - parasitic 659
    - effects 365
      - substrate transistor 415, 544, 556, 620, 656
  - partially depleted collector 81, 87
  - partitioning 617, 635
  - PBSA 583, 584, 628

- PCB model 290
  - PCM 685, 697
  - PDK 22, 26, 634, 686, 695
  - performance optimization 582
  - perimeter effects 334
  - physical effects 434
  - physical limitations 730
  - physics-based compact model 24, 35, 55
  - poly-silicon 106
  - power amplifiers 196
  - power gain 689
  - power transistor 689
  - precision circuits 36
  - predictive modeling 694
  - process control monitor 685
  - process design kit 22, 686
  - process development 54
  - process technologies 681
  - process-independent parameters 682
  - process-specific parameters 635
  - PSD 478
  - punch-through 127
- Q**
- quasi-Fermi potential 214
  - quasi-static 110
    - base current 527
    - transfer current 64, 209, 521, 644
- R**
- RA 58
  - radio-frequency 582
  - random variation 696
  - recombination 235, 434, 483
    - current 114, 452
    - rate 275
    - regional approach 58
    - regional components 59
    - reliability 54, 691
    - RF 582
- S**
- saturation current 111, 114
  - saturation velocity 434
  - saturation voltage 624
  - scalability 24
  - scalable approach 583
  - scalable model parameters 24
  - scaling 303
  - scaling equations 682
  - SCR 59
    - recombination 112
    - width 119
  - second moment 475
  - self-heating 456, 556, 659, 691
  - semi-physical 35
  - series resistances 455, 555, 623, 656
  - SGPM 24, 634, 635, 684
  - SGPM/L0 635
  - shot noise 482
  - SiGe 21, 685
  - SiGe HBT 22, 24, 177, 185, 598
  - SiGe HBT BiCMOS process 687
  - SiGe heterojunction 178
  - SiGe-HBT structure 505
  - single base contact 390
  - single base layout 683
  - single geometry transistor 634
  - size scaling 723
  - small-signal 382, 534
    - model 559

- operation 662, 319
  - SOI HBTs 691
  - space-charge 66
    - density 59
    - region 59
  - S-parameter measurements 588
  - spectral power density 476
  - SPECTRE 25
  - SPICE Gummel-Poon model 24
  - SRH 112
  - standard scaling 335
    - 3D structure 366
  - static characteristics 57
  - static operation 93, 207
  - stationary random process 477
  - stationary stochastic processes 475
  - statistical and predictive modeling 36
  - statistical simulation 686
  - stochastic signal 474
  - storage time 129, 163
  - structural information 55
  - substrate network 545
- T**
- technology 718, 724
    - parameter 685
  - temperature 433
    - dependence 440, 453, 546, 621, 660
  - test structures 54
  - tetrode 587
  - thermal noise 479
  - thermal resistance 610
  - three-dimensional geometry
    - scaling 24
  - THz 717
  - TICCR 274, 275, 276
  - time average 475
  - time-domain 270
  - tolerance 56
  - toolkit 687
  - TP 685
  - TRADICA 583, 635, 684, 696
  - transfer current 93, 208, 257, 337, 449, 549
  - transient integral charge-control relation 274
  - transistor configuration 23
  - transistor dimensions 585
  - transistor sizing 24
  - transit frequency 69
  - transit time 258, 513, 550, 611, 623
  - trap 483
    - assisted tunneling 531
  - tunneling 245, 455, 530
- V**
- variance 475
  - vector network analyzer 588
  - velocity fluctuation 482
  - verilog-A 637, 684, 703
  - verilog-AMS 25
  - vertical dimensions 56
  - vertical non-quasi-static effects 616
  - VNA 588
- W**
- wafer selection 584
  - weak avalanche 116
  - weight function 275
  - Wenzel-Kramer-Brillouin 245

Wiener-Khintchine theorem	X
478	XMOD 687
wireless 25	Z
WKB 245	zero-bias hole charge 446



International Series on Advances in Solid State Electronics and Technology

## COMPACT HIERARCHICAL BIPOLAR TRANSISTOR MODELING WITH HICUM

*Compact Hierarchical Bipolar Transistor Modeling with HICUM* will be of great practical benefit to professionals from the process development, modeling and circuit design community who are interested in the application of bipolar transistors, which include the SiGe:C HBTs fabricated with existing cutting-edge process technology. The book begins with an overview on the different device designs of modern bipolar transistors, along with their relevant operating conditions; while the subsequent chapter on transistor theory is subdivided into a review of mostly classical theories, brought into context with modern technology, and a chapter on advanced theory that is required for understanding modern device designs. This book aims to provide a solid basis for the understanding of modern compact models.